

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:56:08 ; Search time 18 Seconds
(without alignments)
2019.536 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIAESKFKVLNGKSGS.....ATVLKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	5.4	1076	1 RPOB_ASTLO	P27059 astasia lon
2	218	5.4	1630	1 MSP1_PLAFLX	P04932 plasmodium
3	218	5.4	1639	1 MSP1_PLAFLV	P04933 plasmodium
4	215.5	5.4	1251	1 RBP2_PLAFLV	Q00799 plasmodium
5	214	5.3	2339	1 RPL1_PLAFLA	P27625 plasmodium
6	209.5	5.2	1104	1 COLA_CLOPE	P43153 clostridium
7	194	4.8	1957	1 SPOF_SCHPO	Q10411 schizosacch
8	193	4.8	2869	1 RBPI_PLAFLV	Q00798 plasmodium
9	190	4.7	2452	1 RPB1_PLAFLD	P14248 plasmodium
10	186.5	4.6	1790	1 USO1_YEAST	P25386 saccharomyc
11	185	4.6	3259	1 GIAN_HUMAN	Q14789 homo sapien
12	184	4.6	1875	1 MLPI_YEAST	Q02455 saccharomyc
13	183.5	4.6	1053	1 SLPM_BACBR	P06546 bacillus br
14	183	4.5	2334	1 WAPA_BACSU	Q07833 bacillus su
15	179.5	4.5	1183	1 CNA_STRAAU	Q53654 staphylococ
16	178	4.4	1433	1 SUBF_BACSU	P16397 bacillus su
17	178	4.4	1726	1 MSP1_PLAFLP	P04934 plasmodium
18	178	4.4	1726	1 MSP1_PLAFLP	P50495 plasmodium
19	178	4.4	2492	1 ATRX_HUMAN	P46100 homo sapien
20	177	4.4	800	1 CYAA_BACAN	P40136 bacillus an
21	176.5	4.4	1116	1 SLPH_BRECH	P38538 brevivacilli
22	174	4.3	735	1 CIGB_DICDI	Q94481 dictyosteli
23	174	4.3	1176	1 YOH8_YEAST	Q08236 saccharomyc
24	174	4.3	1462	1 GTFD_STRMU	P49331 streptococc
25	174	4.3	1628	1 NAGH_CLOPE	P26831 clostridium
26	173.5	4.3	944	1 NUF1_YEAST	P32380 saccharomyc
27	173	4.3	2710	1 TOXA_CLODI	P16154 clostridium
28	172.5	4.3	596	1 PRIM_CLOAB	P33655 clostridium
29	172.5	4.3	1545	1 IGA3_HAEIN	P45385 haemophilus
30	172	4.3	873	1 SYA_WIGBR	Q8d2w8 wigleswort
31	171	4.2	999	1 HGF2_HAEIN	P44809 haemophilus
32	171	4.2	1701	1 MSP1_PLAFLM	P08569 plasmodium
33	170	4.2	970	1 T3RE_BPP1	P08764 bacterioph

RESULT 1

ID	RPOB_ASTLO	STANDARD;	PRT;	1076 AA.
AC	P27059;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	DNA-directed RNA polymerase beta chain (EC 2.7.7.6).			
GN	RPOB.			
OS	Astasia longa (Euglenophycean alga).			
OG	Chloroplast.			
OC	Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.			
OX	NCBI_TaxID=3037;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=21080522; PubMed=11212895;			
RA	Gockel G., Hachtel W.;			
RT	"Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate Astasia longa.";			
RL	Protist 151:347-351(2000).			
RN	[2]			
RP	PRELIMINARY SEQUENCE OF 1-524 FROM N.A.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=91176556; PubMed=2078869;			
RA	Siemeister G., Buchholz C., Hachtel W.;			
RT	"Genes for ribosomal proteins are retained on the 73 kb DNA from Astasia longa that resembles Euglena chloroplast DNA.";			
RL	Curr. Genet. 18:457-464(1990).			
RN	[3]			
RP	REVISIONS.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=95163124; PubMed=7859309;			
RA	Gockel G., Hachtel W., Baier S., Fliss C., Henke M.;			
RT	"Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate Astasia longa.";			
RL	Curr. Genet. 26:256-262(1994).			
CC	FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.			
CC	CAATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +			
CC	(RNA) (N).			
CC	SUBUNIT: In chloroplasts, the RNA polymerase is composed of four subunits: alpha, beta, beta', and beta".			
CC	SUBCELLULAR LOCATION: Chloroplast.			
CC	SIMILARITY: Belongs to the RNA polymerase beta chain family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			

34	169.5	4.2	1928	1	MYS1_YEAST	P08964 saccharomyc
35	168	4.2	1163	1	SBCC_CLOAB	Q97fk1 clostridium
36	168	4.2	1701	1	MSP1_PLAFLX	P13819 plasmodium
37	167.5	4.2	1541	1	IGA1_HAEIN	P42782 haemophilus
38	167	4.1	1451	1	SPT6_YEAST	P23615 saccharomyc
39	167	4.1	1682	1	MSP1_PLAFL3	P19598 plasmodium
40	167	4.1	3305	1	APLP_WANSE	Q25490 manduca sex
41	166.5	4.1	875	1	ZIP1_YEAST	P31111 saccharomyc
42	166.5	4.1	1147	1	CGA1_HELPY	P80200 helicobacte
43	165.5	4.1	1005	1	RA50_METJA	Q58718 methanococc
44	165.5	4.1	1233	1	SMIA_MOUSE	Q9cu62 mus musculu
45	165	4.1	1176	1	SLAP_BACSH	P38537 bacillus sp

ALIGNMENTS

Db 223 KDVGMEDYIKKNTIENIELIEESKTDKNATKKEE-----KKLYQAOY 274
Qy 76 FTSETGKRMEEYDYKDDRGNIAYDGDLETEKLEIKSIYGVLSPSKDGHFELL 135
Db 275 DLSIYNKOLEE-----AHNLISV-----LE---KRIDTLK-----KNEIKELL 310
Qy 136 GKISNVSKNAKYVGNYSIEIKATKYDFH-----SKMTDFDIYANINDIVDGLAFA 188
Db 311 DKINEIKNPPANGNTPTNLLDKNKKIEHEKEIKETAKTIKFNIDSFTDPL-----364
Qy 189 GDMRLPVKNDOKKAEIKIRMEPIKETKSEYP-----YVSSYGNV-----IEIGE-GDLSK 239
Db 365 -ELEYLRKN-KNIDISAKVETKESTENEPNGVTYPLSYNDINNALNELNSFDLLI- 421
Qy 240 NKPDNLTKMESGIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYTPGKTDML 289
Db 422 -NPFDTYKPSKNIYDNERKKFINEIKIEKIEKKIESDKSYEDRSKSLN-----DIT 476
Qy 290 EG-----NGVYSKE-----DIAKIQVA-----NPNRLALSETTIYADSRNVEDGRSTQSV 334
Db 477 KEYEKLLEIYDSKFNNDLNTNFKRMGRYSYKVEKLTHTNTFASYENSKHNLEKLTJK 536
Qy 335 LMSALDGF--NII-----RQVFTFKMNDGEAIDKGNLVTSSKLVLFKGDKEYT 385
Db 537 ALKTMEDYSLRNIVVEKELKYKLNLSIENEIETLVE--NIKKOEEQLF-----EKKIT 589
Qy 386 GBDKFNVEAIKEDGSMFLDTDKPVLNLSMDKNYFNPSKSNKIYVRNPEFYLRGKISDKGOF 445
Db 590 KDNKPDEKILEVSDIVKVOQVLL--MNKIDELKTKQLILKNVE--LKHNIHVPSY 644
Qy 446 NMLRVNESVVDNLIYLGOLHIDNTDFNIKU-----NVK-----480
Db 645 KOENKQE----PYLLIVLKEIDKLKVPKPVESLINEEKKNIKTEGQSDNSEPSTGEI 700
Qy 481 -----DGDIMDMGMDYK-----ANGFPDKV 501
Db 701 TGOATTKPQOQAGSALLEGDSVOAQOQOAPVPVPEAKAQVPTTAPVNNKNTNV 760
Qy 502 TMDGNYVLTQYSOLNA-----KAVGVHYQFLYNDVKNPEVNDPKGNTSIEYADGKSV 556
Db 761 SKLD---YLEKLYEFLNTSVYCHKYLVSHSTWNEKILKQYKITEEESKLSGCPDLIL 817
Qy 557 FNIN-----DKRNGPDG---EIQEQHIYNGKEYTSFNDIKQIIDKTANIKIV 603
Db 818 FNIQNNIPVMSFMSLNNLSQLPFMEIYKEMVNCNLYKLDKNDKIKNLEBAKKVSTSV 877
Qy 604 KDPARTTVKEPILNKOTGEVSELKPHRVTVTIQNGKENSSTIVSEEDFILPVYGELEK 663
Db 878 KTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLNLSKLFLNLS-----LGKNKN 928
Qy 664 GYQFDGWEISGPEGK-----KDAGYVINLSKDTFIKPVFKTIEEKEENKPTF--D 713
Db 929 IVQ-----ELIQKSENFEKILKQSDTYNFSFTNFKSKADDNLSNDESKRKLLED 984
Qy 714 VSKKDNPNQVNSQLNSHREKDLQREHSQKSDSTKVTATVLDKNINISSK-STNNP 771
Db 985 INKLKKTQLSFLDLYNKYKLERLFDKKTKGVKQWKQIKLTLKLEQLESKLSLNNP 1043

RESULT 3
MSPI PLAFW
ID MSPI PLAFW STANDARD; PRT; 1639 AA.
AC P04913;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (p195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]

SEQUENCE FROM N.A.
MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273 (1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; ICEJ; 28-MAY-99.
DR InterPro; IPR006209; EGF like.
DR Pfam; PF0008; EGF; 1.
DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1639
FT CARBOHYD 116 116
FT CARBOHYD 268 268
FT CARBOHYD 764 764
FT CARBOHYD 768 768
FT CARBOHYD 783 783
FT CARBOHYD 844 844
FT CARBOHYD 920 920
FT CARBOHYD 964 964
FT CARBOHYD 1058 1058
FT CARBOHYD 1165 1165
FT CARBOHYD 1174 1174
FT CARBOHYD 1445 1445
FT CARBOHYD 1526 1526
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
Query Match 5.4%; Score 218; DB 1; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.025;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
Qy 23 KDTTC-VEHHQENESIK-----EKSTFTDRNISTIRDENKDLKLIKKEFREVD 75
Db 232 KDVGMEDYIKKNTIENIELIEESKTDKNATKKEE-----KKLYQAOY 283
Qy 76 FTSETGKRMEEYDYKDDRGNIAYDGDLETEKLEIKSIYGVLSPSKDGHFELL 135
Db 284 DLSIYNKOLEE-----AHNLISV-----LE---KRIDTLK-----KNEIKELL 319
Qy 136 GKISNVSKNAKYVGNYSIEIKATKYDFH-----SKMTDFDIYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGNTPTNLLDKNKKIEHEKEIKETAKTIKFNIDSFTDPL-----373
Qy 189 GDMRLPVKNDOKKAEIKIRMEPIKETKSEYP-----YVSSYGNV-----IEIGE-GDLSK 239
Db 374 -ELEYLRKN-KNIDISAKVETKESTENEPNGVTYPLSYNDINNALNELNSFDLLI- 430
Qy 240 NKPDNLTKMESGIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYTPGKTDML 289

Db 1058 TIISKUKGVIIWENTMTTSSAKETALYNELKNKTSLSNIYQTSNEVKLOEMKS 1117

Qy 745 KSDSTKDVTA---TVLD---KNNISSKSTNN 770

Db 1118 NADKYIDVSKIFNTVLDTKQSNIVTQHINN 1149

RESULT 5

RPC1 PLAF A

ID RPC1 PLAF A STANDARD; PRT; 2339 AA.

AC P27625;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92018020; PubMed=1656254;

RA Li W.B., Bizik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;

RT "Characterization of the gene encoding the largest subunit of Plasmodium falciparum RNA polymerase III.";

RL Mol. Biochem. Parasitol. 46:229-240(1991).

CC -! FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.

CC -! CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).

CC -! SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE III.

CC -! SUBCELLULAR LOCATION: Nuclear.

CC -! MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S AND TRNA GENES.

CC -! SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M73770; AAA29729.1; -

DR InterPro; IPR007072; RNA_pol A.

DR InterPro; IPR007080; RNA_pol_Rpb1_1.

DR InterPro; IPR007066; RNA_pol_Rpb1_3.

DR InterPro; IPR007083; RNA_pol_Rpb1_4.

DR InterPro; IPR007081; RNA_pol_Rpb1_5.

DR InterPro; IPR006592; RNA_POLA_N.

DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.

DR Pfam; PF06623; RNA_pol_Rpb1_2; 1.

DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.

DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.

DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.

DR SMART; SM00663; RPOLA_N; 1.

KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;

FT ZN FING 88 101

SQ SEQUENCE 2339 AA; 272829 MW; EDD899363086FD48 CRC64;

Query Match

Best Local Similarity 21.6%; Score 214; DB 1; Length 2339;

Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

Qy 24 DTGVEHHQOE--NEESIEKESFTIDRNISTIRDF---ENKDLKLIKKKPREVDDFTS 78

Db 1292 DNTYVEQIEWKELSKNKTKEQSFK-----GTIDRMHEDSEQMNKFTTKAK----FPI 1342

Qy 79 ETGRMEBYDYKDDKGNIIAYDDGDTLEYETEKEDEIKSKIY-----G 122

Db 1343 EK-KKGKHECNDDEIYNTQYD---NIQYNNISNYIKSQNLNTHQVNDLSPKNN 1398

Qy 123 VLSPSKDGH-----FEILGKISNVSKNAKYVYGNKYKSIETATKYDFHSKMTWPD 174

Db 1399 VILPPKEYHSIFHFVNDYRNVEIKNLMDKKKIFLNSEKN---VQSKYNNRMSKDKKKI 1456

Qy 175 YANINDIVDG-----LAPAGDMRLFVKNDQ--KKAETIKRMPEK---IKETKSE 219

Db 1457 EI-INNIYRNEKKLNRWKTMDNDNTWSSDDSDSIIAKKIIKKNKEKKYHPKEKEN 1515

Qy 220 YPVSSYGNVIELGEGDLSKNKPDNLTK-MESGKIYSDSEKQOYLKKNIIILKGVYALK 278

Db 1516 PDR-NNYKMITDNNNNNDNNNNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1569

Qy 279 TTYNPGKTDML--EGNGUYYSKEDIAKIQKANPNL-----RALSETIIVADS-- 322

Db 1570 TNYN---TNIYPNDCNGIYEKETNNNELTNSNMCDKNDFSEDFNNINENDLIDNKY 1626

Qy 323 -----RNVEDGRSTQSVLMSALDGFNIIRYQVF---TFKMDKGBAIDKDGMLVDSSKL 374

Db 1627 YRQIFKNVIGFVSVEYVESYKQHYLPYELIKWTSLELLEYLTILPTNIFLHTKLSK- 1685

Qy 375 VLFQGD---DKYEYTGEDKFNVEAIKEDGSMFLFDTPKPNVLSMDKNYFNSKNKIYVRNP 431

Db 1686 ---KEKPTHQKNTGKMKIYBEIKK---WLFI---KAINI---YKYFSPKKSIEL---- 1728

Qy 432 EYFLRGKISDKGGFNWELRVNESVDVNLIIYGLDLDHIDTRD-----FNI-----KLVKD 481

Db 1729 -----IKKKDYFNYIHK-NYDISHRYIIH-DYSFINLKQLYLFIFNFIYKPKYISTP 1779

Qy 482 GDIMDWGMKDYKANGFPDKVTDMDCNVYLQTCYSDLNAKAVGVHYQVFLYDVKVPEVIDP 541

Db 1780 GD-----AVGSIQAQSIGEPGTQMTLTKTHFAGVAMNV-TLGV-----PRIKEIINASN 1828

Qy 542 KNTSIEYADGKSVVFNINDKENNGF-----DGEIQEQTIIYINGKEYTS-----F 586

Db 1829 SIQTPI---LNIPLVDNDNYPALMMKSKLEKTTIRDICMYIK-EDYTSRGVFLSVKF 1882

Qy 587 ND-----IKQIDKTLNI-----KI---VVKDFARTTVK--EFLNKOT 621

Db 1893 NEELIQKFLNINAYNIKDIILKQSHINKINKINKINKINKINKINKINKINKINKINK 1938

Qy 622 GVSLSKPHRVTVTTIONGKEMSSITVSEDFILPVYKGELEKGYQFDGWEISGFEGK--K 679

Db 1939 FOMESLKGGLDLLIYGDYKDKRCIIKKEDI-----EVTNDEDEICD 1980

Qy 680 DAGYVINLSKDTFIKPVFKIIEBK---EENKPTFDVSKKDNFQVNVHSQLNESHKED 736

Db 1981 DMDEYNNVSQGT-----ELVERKCSKEENKNAIRVKKEID-----DN 2019

Qy 737 LQREB-----HSQKSDSTKDVATVLDKNNISSKSTNN 770

Db 2020 LEKEENIIYVSEKDSVNLKSEKKDIN---DDNNNDNDDNNNN 2060

RESULT 6

COLA_CLOPE

ID COLA_CLOPE STANDARD; PRT; 1104 AA.

AC P43153;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).

GN COLA OR CPE0173.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.

RC STRAIN=NCIB 10662 / Type C;
RX MEDLINE=94110220; PubMed=8282691;
RA Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;
RT "Purification and characterization of Clostridium perfringens 120-
RT kilodalton collagenase and nucleotide sequence of the corresponding
RT gene.";
RL J. Bacteriol. 176:149-156(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [3]
RP SEQUENCE OF 1073-1104 FROM N.A.
RC STRAIN=NCIB 10662 / Type C;
RA Matsushita O.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Digestion of native collagen in the triple
CC helical region at Xaa-[Gly bonds. With synthetic peptides, a
CC preference is shown for Gly at P3 and P1'; Pro and Ala at P2 and
CC P2'; and hydroxyproline, Ala or Arg at P3'.
CC -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.
CC -!- SIMILARITY: Contains 1 PKD domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13791; BAA02941.1; ..
CC EMBL; AF003185; BAB79879.1; ..
CC EMBL; D50309; BAA08848.1; ..
CC PIR; A36866; A36866.
CC MEROPS; M09.002;
CC InterPro; IPR002169; Miccollptase.
CC InterPro; IPR000601; PKD domain.
CC InterPro; IPR006025; Zn_Mrpptdase.
CC Pfam; PF01752; Peptidase_M9; 1.
CC Pfam; PF00801; PKD; 1.
CC Pfam; PF04151; PPC; 2.
CC PRINTS; PR00931; MICOLLPTASE.
CC SMART; SM00089; PKD; 1.
CC PROSITE; PS50093; PKD; 1.
CC PROSITE; PS00142; ZINC PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 86
FT CHAIN 87 1104 MICROBIAL COLLAGENASE.
FT DOMAIN 774 862 PKD.
FT METAL 502 502 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 503 503 BY SIMILARITY.
FT METAL 506 506 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 38 38 L -> F (IN REF. 1).
FT CONFLICT 722 722 I -> M (IN REF. 1).
FT CONFLICT 748 748 G -> E (IN REF. 1).
FT CONFLICT 945 945 V -> E (IN REF. 1).
FT CONFLICT 970 970 T -> A (IN REF. 1).
FT CONFLICT 987 987 A -> E (IN REF. 1).
FT CONFLICT 1098 1098 I -> T (IN REF. 1 AND 3).
SQ SEQUENCE 1104 AA; 125935 MW; F4B7377194ED021C CRC64;

Query Match 5.2%; Score 209.5; DB 1; Length 1104;
Best Local Similarity 20.0%; Pred. No. 0.036;
Matches 194; Conservative 136; Mismatches 345; Indels 295; Gaps 50;

QY 6 AESKFNGLNGKESGLKDDTTGVEHHQENBESIKESSFTIDRNISTIRDFENK----- 60
DB 40 ADKVENSNNGQ---ITREINADQISKTELNEVATDNNRPLGSPISAPSRARNKIYTPD 96
QY 61 -----DLKKLIK-KPFEVDD-FTSETGK-----RMEFYDYKYDKGNIIAYDD 102
DB 97 ELNRMNSYDLVELIKTISVENVPDLFNFNDGSYTFFSNRDRVQAIYIGLEDSRTYADD 156
QY 103 GTDL-----EYETEKLDEIKSIY-----VLSPSKDGHF 132
DB 157 DXGIFLVEFLRAGYLLGYFNKQLSYLNTPLQKNECLPAMKAIQNSNFRUGTKAQGVV 216
QY 133 ELIGK- ISNVSKNAKY-----YGNNY-----KSIKATKYDF 165
DB 217 EALGRLGNASADPEVINNCIYVLSDFKDNIDKYSNYSKGNVFNLMKGIYINSYIY 276
QY 166 HSK---TMTFDLYANI-----ND-----IVDGLAFAGDMRLFVKDNDQ 200
DB 277 NTKGYDAKNTEFYNRIDPYMERLESCTIGDKLNDNNAWLVNVALYITGRMGKFPEDPSI 336
QY 201 KKAEBIKIRMEPEKIKETSEYPYVS-----SYGNVIELGEGDLKSNKPDNLTKMESGKIYSD 256
DB 337 SQALERAM-----KEYPILSYQYIEAANDLNFQ--CKNSSGN--DIDFNKIKAD 384
QY 257 SEKQOYLLKDNIIILRKGYALKVTTYNPGKTDMLGNGVYSKEDIKIAKQANPNLRALSET 316
DB 385 A-----REKYLPTYTFDDGKFVVKAGDKV-TEBKIKELYASKEVKAQFNR 430
QY 317 TIYADSRNVEDGRSQSVLMSALDGFNIIRYOVFTFKMND--KGEAIDKQGNLVTDSKL 374
DB 431 VYQND- KALEGNPDILTV-----VIYNSPEYKLNRIINGFSTONGYIYENIGTF 482
QY 375 VLFQKDDKE--YTGEDKNVE-----AIKEDGSMFLIDTKPVNL--- 411
DB 483 FTYERTPEESIYTLBELFRHEFTHYLQGRYVVGWVGGEFYQEGVLTYEESGTAEFFAG 542
QY 412 SMDKNVFNPSKS---NKIYVRNPEFYLRGKISDKGGFNWELRVNVSVDNLIYIGDLHID 468
DB 543 STRTDGIRKPSVTOGLAYDRNRMSLYGLVHLAKYG-SWDFYNYGFALSNY-----MYNN 596
QY 469 NTRDFNIKLN-VKDGIMDMWKMOKYKAN-----GFPDKVTD-MDGNVYLQTYSDLNKA 521
DB 597 NMGEMKNMTNLIKNNDSV--GYKDIASMSDYGLNDKYQDYMS---LLANNIDNLDPVL 651
QY 522 VGVHY-----QFLYDNVKEVNI-DPKQNTSIEYADGKSVFVNINDKNNNGFDGRI 571
DB 652 VSDEYVNGHEAKDINEITNDIKEVSNIKDLSNVE-----KSQFTTYDMRGTT----- 699
QY 572 QEQHIYINGKEYTSFNDIKQIIDKTLNLIKIVVKDFAR-----NTTVKEFILNKDTGE-- 623
DB 700 -----YVGRSOGCEENDWK---DMNSKLNILKELSKSWNGYKVTATYFVNHKVDGNGN 751
QY 624 -VSELKPHRV-----TWTIONKEMSTIVSEEDFILPVYKBLEKGYQPDGWEISGFEGK 678
DB 752 YVVDVVFHGMNTDTNTDVHVNKEPKAVIKSDSVI-----VEBEINFDGTESKDEDEGE 804
QY 679 -----KDAYVINLSK-----DTFIKPVFKLEEKKER- 706
DB 805 IKAYEWDFGDGSKSNEAKATHKYNKTYGEYEVKLTVDNNGNGINTESKKT-KVEDKPEVE 863
QY 707 --ENKPTFDVSKKKNQPNVHNSQLNESHK--KEDIQREHSHSQKSDSTKDVATATVLDKNNI 762
DB 864 INESEPNDFEKA-----NQIAKSNMLVKGTLSEEDYSDK-----YYFDVAKKGNV 909
QY 763 SSKSTNNPN 772
DB 910 --KITLNNLN 917

RESULT 7
SPOF SCHPO
ID SPOF SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; Q9USE9;

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SPO15 OR SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
RA Ikenoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
RT to the spindle pole body and essential for its modification";
RL J. Cell Sci. 113:545-554(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 705-871 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: Has a role in the initiation of spore membrane
CC formation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z70690; CAA94624.1; --
CC EMBL: AB027811; BAA87115.1; --
CC PIR: T38077; T38077.
CC GeneDB_Spombe; SPAC1F3.06c; --

RESULT 8

KW Sporulation; Coiled coil.
FT DOMAIN 199 785 COILED COIL (POTENTIAL).
FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
Query Match 4.8%; Score 194; DB 1; Length 1957;
Best Local Similarity 20.5%; Pred. No. 0.33;
Matches 174; Conservative 147; Mismatches 290; Indels 238; Gaps 44;
QY 4 ETAESEKFNGLNGKESGLKQDTTGVHHQENEE-----SIKEKSFIDRNISTY 54
DB 937 KIEESKSDLG-----KKLTARQEEISLNKEENMSQAITSVKSLDETLSSKSLE 999
QY 55 RFENKDLKLLIKKIFREVDFTSETGKRMEDYKYDD---KGNIIAYDDGDTLEETE 111
DB 990 ADIEH-----LNKVSEVE---VERNALLASNERLMDDLKNGENIA-SLQTEIEKKRA 1039
QY 112 KLDEIKSKIYGVLSPSKDGHFELGKISNVSKAKVYVYNNYKSIIEIKATKYDFHKTWT 171
DB 1040 ENDDIQSK-----LSVVSSEYENLLIISQTNKSLDKTNQLKYIEK--- 1081
QY 172 FLYIANINDVGLAFAGDMRLFVKDNDQKAEIKIRPEKIKETKSEYPYVSSVGNVIE 231
DB 1082 -----NVQKLLD-----EKQORNVLE-----ELTSKYGKL-- 1107
QY 232 LGEGLSKNKPDLTKMESGKIYSDSEKQYLLKDNII--LRKGVALKVTTVNPNGTMDL 289
DB 1108 ---GEENAIKDELLALK-----KSKQHDLCANFVDLKE-----KSDAL 1146
QY 290 EGNVYSKEDIKIQANPNLALSE-----TTIYADSRNVEDGRSTQVLSMALDGNF 343
DB 1147 EQLTNEKNELIVSLQSNNEALVEERSDLANRLSDMKKSLSDSNVIRSIDL--- 1202
QY 344 IIRYQVFFKMKDKEADKQGNLVTDSKLVFGKDD--KYTG-EDKFNVEAKEDQS 400
DB 1203 -----VRVNDLDTLKKDKDSLSTQYSEVQCDRDDLLSLKGEESFNKYAV---S 1250
QY 401 MLFIDTK-----PVNLMDKNYF---NPSKSNKIYVRNPEFYLRGKISDKGFGNWLVR 451
DB 1251 LRELCTKSEIDVPVSEIILDDNFVFNAGNFSELSRLTVLSLENYL-----DAFN-QVNF 1302
QY 452 NESVVDNYLIYGLDHIDTRDFNIKLN-----VKDGDIMDMKDYKANGFPDPKVTDM 504
DB 1303 KMWELDNRLTTTDAEFTKVADLEKQLQHEHDDWLIQRGD-LEKALKDSEKN-FLRKEAM 1360
QY 505 DGNVY-LOTGYSDLNKAVGVHYQFLYDNV-----KPEVNIDPKGNTSIYADG-----KS 554
DB 1361 TENIHSLEEGKEE-TKGIAELSLSRLEDNQLATNKLKNQLDHL-NOEIRLKEDVLKES 1418
QY 555 VYFNINDKRNNGFDGEIOBOHIYINGKEYTSFNDIKQIIDKT--LNKIVVKDFARNTTV 612
DB 1419 LIISLEESLNN-----QROKESLLDAK-----NELEHMDTSTRKNSSLMKEKIESNSSL 1469
QY 613 --KEFILNKDTGEVSEL-KPHRVTVT-----IQNGKE---MSSTIVSEEDFLLPVYK 658
DB 1470 DDKSFELASAVEKLGALQKLHSELSLMENIKSQLQEAKEKIQVDSTIQLDHEITASK 1529
QY 659 GELEKGYQFDGWEISGFGKDKDAGVIVNLSKDTFTKPVFKKTEE----- 702
DB 1530 -----NNYEGKLN-----DKDSIIRLSENIEQLNLLAEKSAVKRLS 1568
QY 703 -KKEE-----NKPTFDVS--KKDNQPNVHNSHREKDLQREHSQKSDTK--DVT 753
DB 1569 TEKESIEQLFNSRLADLEYHKSQVSELSRQVSELSRQVSELSRQVSELSRQVSELSRQV 1628
QY 754 ATVLDKNNI 762
DB 1629 NOVKDLSNI 1637

BBP1_PLAVB STANDARD; PRT; 2869 AA.

AC Q00798; 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Reticulocyte binding protein 1 precursor.

GN BBP1.

OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=31273;

RN [1] SEQUENCE FROM N.A.

RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites."

RT Cell 69:1213-1226(1992).

CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO HUMAN RETICULOCYTE CELLS.

CC -!- SUBUNIT: Homodimer (Potential).

CC -!- SUBCELLULAR LOCATION: Membrane-bound.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M80897; AAA29743.1; --

KW Malaria; Receptor; Signal; Transmembrane.

FT SIGNAL 1 17 POTENTIAL.

FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.

FT DOMAIN 18 2807 EXTRACELLULAR.

FT TRANSMEM 2808 2826 POTENTIAL.

FT DOMAIN 2827 2869 CYTOPLASMIC.

FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).

FT SEQUENCE 2869 AA; 330213 MW; B9DBB442205BECFF CRC64;

Query Match 4.8%; Score 193; DB 1; Length 2869;

Best Local Similarity 20.2%; Pred. No. 0.58;

Matches 204; Conservative 153; Mismatches 329; Indels 322; Gaps 56;

QY 1 KLGETAESKFNILG-----NGKEGSL---KKDTTGVGH-----HHQF-----N 35

DB 1697 KVKAARESEYKNETVQNEMSRINVEGSLTDIDKTTIDNDLLQMKQYEEGLLQKIK 1756

QY 36 EESIKEKSSP-TIDRNIISTIRPENKDLKLKKKPREVDDFTSETGKMEEDYKYDKK 94

DB 1757 ENADKRSNFELVGSSEINALLDPTSGTIFIKLKEYDMTGDL-KNYGVKMEIHGEFTKS 1815

QY 95 GNIIA--YDDGTDTLETETK-----LDE 115

DB 1816 YNLIETHLSNATDYSVTFEKAQSLRLAEKBEHLRRREBEALFLNDIKKVESLKLKE 1875

QY 116 IKSKIYGVLSPKDGHFEI-----LGKISNVSKNAKYVYGNKYSI--EIRA 160

DB 1876 MKKVSAAEYEGMKRDTSVSOLVQDMKTTIVDELKTLNDISECSVL--NNVSVIVKKVE 1933

QY 161 TKY-----DFHSKTMFTDIYAN--IND--IVDGLAFAGMDLRFVK----- 196

DB 1934 SKHADYRRDANSYMEGSMVTLYANFLSDEAKISSGMEFNAEMKSNFNTDLELIFSVISNS 1993

QY 197 -----DND--OKKAE-----IKIRMPKIKETKSEYPIYVS--- 225

DB 1994 NELLKKIQDSNDVIQKERESEQLAKDADTIYVNIKLKNEFKLEAKNKEVSEKVR 2053

QY 226 -----YGVNIELGEGDLSKNKPNLTVMESGKIYSD--SEKQOYLK-DN 267

DB 2054 EALKRLSQVEGIRCHFNFRHLDD---NTEELNLAQKWT--IYRDKKSERESGLQEMEN 2108

QY 268 IILRGYALKVTTYNPGTKDMLG---NGVYSKEDIKIQKAPNLRALSETTIIVADSRN 324

DB 2109 -----EMTYNSNIT-QLEGIVSAGESKEDIEKLERNEEMRNISEKISTDSKV 2158

QY 325 VEDGRSTOSVMSALDGFNIIRYQVTFKMDKGEA-----IDKGNLVTDSSKLVLFGK 379

DB 2159 IE-----VNSTIDEL-----YKLGKNCQAHWISLISYTANMKT-SKKLIMINK 2200

QY 380 DDK-----EYTGDEKFNVEAIKED-----GMLFI-----DTKPNLSM-DKNY 417

DB 2201 EKENTEKCDVYIKDSSSSDTGCVETLKGYPGSKLTFSSASEIVONADTYSVNFAKHEKS 2260

QY 418 FNPSSNKIYVENPEFYLRGKISDKGFWELRVNESVDNYL-IVGDLHID----- 468

DB 2261 LNAIRDIK-----KELYLFHQNSD-----ISIVEGGVQNMLALYDKLNEKREMDLEY 2308

QY 469 -NTRDFNFKLVNKGDI-----MDWGM-----KDYKANGFPDPKVTMDGNNVLIQ 511

DB 2309 RNISETKLQMEHSTDFPKPMIELHKGWNETNKSILLEKEKLVNDHMSMEAM-IK 2367

QY 512 TG--YSDLNAKAVGVHYOFLYDNVPEVNIIDPKGNTSIYADGKSV-----FNINDKR 563

DB 2368 NGLKYPSPSVQNNIYSVIEAEVKTLEIDR-----DYGDNYQIVEEHKKQFSILIDR 2421

QY 564 NNGFDGEIQ---EQHIY---INGKEYTSFND-IKQIDKTLNLIKLVKDFARNTTVKEF 615

DB 2422 TWALMDDIEFKKENYNLMVNTIHRVNDYIEKINKLVQAKTEYEQILEN--IKQ- 2478

QY 616 ILNKDTGEVSELKPHRVVTIQNGKEMSGSTIVSEEDFILPVYKGE--LEKGYQDFGWEIS 673

DB 2479 -NDDMLQNLFLKVSIIIEYPENVVKKESILND-----LYEQERLLKIGEHLD----- 2525

QY 674 GREGKQAGYVNLNKDTPIKVPFKKIEBKEENKPTFDVSKKQNDKPNVNSHLSHR 733

DB 2526 --EIKRNVTTLS-----SYEIDQKMEMMSKNLLE-----KSKMNTYSIELER 2569

QY 734 K-EDIQREHSOKSDST--KDYTAIVLDKNN-----ISSKSTNNPNK 773

DB 2570 EANEINRDAQIKDDDDTILNSVLEAAIKRGMDAIFQMSADRNPN 2617

RESULT 9

RP1_PLAVD

ID RP1_PLAVD STANDARD; PRT; 2452 AA.

AC Pl4248;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).

GN RP11.

OS Plasmodium falciparum (isolate CDC / Honduras).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5836;

RN [1] SEQUENCE FROM N.A.

RX MEDLINE=9009832; PubMed=2690004;

RA Li W.B., Bizik D.J., Gu H., Tanaka M., Fox B.A., Inselburg J.;

RT "An enlarged largest subunit of Plasmodium falciparum RNA polymerase II defines conserved and variable RNA polymerase domains."

RL Nucleic Acids Res. 17:9621-9636(1989).

CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

CC -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO TWELVE DIFFERENT POLYPEPTIDES. THIS POLYPEPTIDE IS THE LARGEST COMPONENT OF RNA POLYMERASE II.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- PTM: THE TANDUM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED. THE PHOSPHORYLATION ACTIVATES POL2.

CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE

RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RP [3]
 CC SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC !- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC !- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC !- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC !- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54378; CAA38253.1; -
 CC EMBL; L03188; AAB00143.1; -
 CC EMBL; US3668; AAB66659.1; -
 CC SGD; S0002216; USO1.
 CC InterPro; IPR002017; Spectrin.
 CC InterPro; IPR006955; Uso1_p115_C.
 CC InterPro; IPR006953; Uso1_p115_head.
 CC Pfam; PF04871; Uso1_p115_C; 1.
 CC Pfam; PF04869; Uso1_p115_head; 1.
 CC TransPort; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 CC DOMAIN 1 724 GLOBULAR HEAD.
 CC FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 CC FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 CC FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 CC FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 CC FT CONFLICT 847 847 G -> E (IN REF. 2).
 CC FT CONFLICT 924 924 E -> K (IN REF. 2).
 CC FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 CC FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 CC FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 CC FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 CC FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 CC FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 CC FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 CC SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
 CC
 CC Query Match 4.6%; Score 186.5; DB 1; Length 1790;
 CC Best Local Similarity 18.1%; Pred. No. 0.63;
 CC Matches 156; Conservative 158; Mismatches 277; Indels 269; Gaps 41;
 CC
 CC QY 1 KLGEIAESK-----PKNLNGKEGSIKKDTTGVHHQHNSIK-----EKSFF 45
 CC DB 1076 KISELTKTRELEAEALAAANKLELETLETSEKALK-EVNEEHLKEEKLQLEKAT 1134
 CC
 CC QY 46 TIDRNISTIR-----DFENKDKLKLKKFREVDFTSETGKRMEYDYKYDDKGNIIA 99
 CC DB 1135 ETQQLNSLRANLESLEKEHDLAAQLK-----YEQIANKEQYNEEISQLN---- 1183
 CC
 CC QY 100 YDGTDLLEYTE-----KLDEIKSIYGVLSPS-----KDGHPFIIG-----KISNVSK 143
 CC DB 1184 -DEITSTQOENESIKKNDLEGEVAMKSTSEQSNLKKSEIDALNLQIKELKKKNETN 1242
 CC
 CC QY 144 NAKVYGVNNVKSIEIKATKY-----DFHSKTMFTDLYANINDIVDGLAFAGDMELFVK 196
 CC DB 1243 EASLL--ESIKSVESSTVKIKELQDCNFKEK-----EVSEEDKLKASDDKNSYL 1292
 CC
 CC QY 197 DNDOKKAEIKIRPEKIKETKSYPPYVSSGVNIELGEDLS-----KNKPDNLT 246
 CC DB 1293 ELQKSEKIKEELDAXTELKIQLEKITNLAKAKESELSRLKKTTSBERKNAEQLE 1352

QY 247 KMSG-KIYSDS-EKQVLLKD-NILRKGVALKVTYTPNGKTDMLGNGVYSKE----- 298
 DB 1353 KUKNEIQKNQAFERKLLNNGSSTITTOYSEKINTLEDELIRIQNELKAKELDNT 1412
 QY 299 -DIAKIQANPNLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVFTFMNDK 357
 DB 1413 SELEKVSLSNDELLEEKQNTI---KSLQD-----EILSKDKITRNDK 1453
 QY 358 GRAIDKDGMLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLFDTPVNLSDKNY 417
 DB 1454 LLSIERDNKRDLSESLKEQLRAAQESKAKVEE--GLKKLEESS-----KEKAELEKSKEM 1506
 QY 418 FNPSSNKIYVNPFPYLRGKISDKGFNWLVRNESVVDNLYLYGDLHIDNTR---DF 473
 DB 1507 MKKLEST---ISNETELKSSME-----TIKSEDEKLEQSKSAEEDIKULOHEKSDL 1556
 QY 474 NIKLVNKGIDIMDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYD 533
 DB 1557 ISRINESEKI-----EEL 1570
 QY 534 KPEVNDPKGNTSIEYADGKSVVFNINDKRNNGFDGEIIOEHYIYNGKEYT-----SFNDI 589
 DB 1571 KSKLRIEAKSGSELE-----TVKQELNNA-----QEKIRINAENTVLKSKLEDI 1615
 QY 590 -KQIIDKTLNIIKIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSST--- 645
 DB 1616 ERELKDKQAEIK-----SNOEKELL-----TSRLKELE-----QELDSTQOK 1653
 QY 646 -IVSEEDFILPVYKGELEKGYQFD-----GWEISGFEGKKDAGYVIN 696
 DB 1654 AQKSEERRAEVRKQVEKS-QLDKAMLETKYNDLVNKEQAW-----KRD----- 1699
 QY 687 LSKDTFKVPFKIEKEEENKPTFDVSKKDKQNVNHSQLNESHKEDLDQREHSQKS 746
 DB 1700 --EDT---VKTTDSQOEIEK---LAKELDLNKAENSKLKEAN-----EDRSEID 1742
 QY 747 DSTKDVATVLDKNLSSKS 766
 DB 1743 DLM--LLVTDLDKNAKYRS 1760
 RESULT 11
 ID GIAN HUMAN STANDARD; PRT; 3259 AA.
 AC Q14789; Q14398;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).
 GN GOLGB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94187728; PubMed=7511208;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin).";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94257116; PubMed=8198703;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).

CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -!- SIMILARITY: Contains 3 s-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M15364; AAA87321.1; --
CC EMBL: M14238; AAA22372.1; --
CC EMBL: M19115; AAA22760.1; --
CC EMBL: M31828; AAA22619.1; --
CC PIR: A28555; A28555.
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH; 2.
CC PROSITE: PS01072; SLH DOMAIN; 2.
CC Cell wall, S-layer; Signal; Repeat.
CC KW SIGNAL 1 23
CC FT CHAIN 24 1053 MIDDLE CELL WALL PROTEIN.
CC FT DOMAIN 29 92 SLH 1.
CC FT DOMAIN 93 143 SLH 2.
CC FT DOMAIN 144 203 SLH 3.
CC FT SEQUENCE 1053 AA; 117146 MW; DB4213188D9D5E4F CRC64;
CC SQ

Query Match 4.6%; Score 183.5; DB 1; Length 1053;
Best Local Similarity 20.2%; Pred. No. 0.45;
Matches 194; Conservative 143; Mismatches 326; Indels 297; Gaps 52;
QY 10 FRLNGKRGSLKQDTT-GVEHHQNEESIKESFTIDRNISTIRDFE-----NK 60
DB 183 FRLMDNALVDLMEQVEFGTDIRHEITKTLLTKYLKVV-----TVRDMWAQAGNDSE 236
QY 61 DLK-----KLIKKKFREV-----DDFTSETGKMEYDYKY-----DDKG 95
DB 237 DPLVLTNPVPAIGLKIKANEVTLNGKDGAGINTTYKVADGINANDFGQHVQVWIKDDKE 296
QY 96 NIIAYDDG-TLEYETEKLDE--IKSKIYGLVSPSKDGHFEILGKISNVSKAKVYVGN 152
DB 297 DIVVMWEGTDQVIMDRVGEFTLKGKT-----FEDPKDLSN-SDLADL----- 339
QY 153 YKSIIEIKATK--YDFHSKWTMTFDLYANINDIVIDGL-----AFAGDMRLFKVDNQ 200
DB 340 --KLELDASEKSYRKNKTKVYNTFRNDFNDPVDGLKEILKNADGCGFTGAKV-VLDNN 396
QY 201 KKAETIKRMPEKIKETKSEYVPVSSYGVNVELGE-----GDLS----- 238
DB 397 ETAYIHVIDDQSMNKEDEGVKYGSEVISKIDTKKKITNRDNDKFDLDGKEGKDFLVF 456
QY 239 -KNKPDNLTKMESGKIYS-----DSEKQOYLKONILKGVALKVVTYVNPCKTDLMBG 291
DB 457 LNKPAKPSDLKEGMWISYVYADGDDKLLVFATDTV--EGKVDKVRNNNDVRLTIG 514
QY 292 NGVY-----SKEDIAIKQANPNL-RALSETTI--YADS-----RNVE-----DG 328
DB 515 DKTYRYEGATSDGDKNDQVQIDKDHDLVDSDDTEVKLYLDASGRVPHIETKDAIDD 574
QY 329 RTQSVLMSALDGFNIIRVQVFTFK-MNDKGBAI-----DKDGNLVTDSKLVLF 377
DB 575 RK-QRAIVTRSATNTSK-DTWDFRVLTKQKEITVSLKAKNIYDPDGKNFSDRNK----- 628
QY 378 GKDDKE---YTGEDK-----FNVEAIKEDGSMFLDTKPNLSMDKN----- 416
DB 629 NQDDLEDILVPSKDKDTLLVLTLDADGPKGVKEFL--KPVKVEQESGKAWDDLADDD 686
QY 417 -----YFN-----PKSKNIYVRNPEFLYLRKISDKGPNFELRVNES----- 454
DB 687 MVGDYEVTDKTAVFNTWTKLESSKREKELKNAKTAFAKDVADENDLSVIYTVNDKDEVEA 746
QY 455 --VVDNYLYGLDHLIDNTRDF-----NIKLVNKDGD-IMDWGMKDYKANGFPDKVTDM 504

DB 747 IFVFGDGLTGDAHYGVQVIDFGRKGGKOTIRVWEKDGDKVBE---KEYLDDGDDQDLKDE 803
QY 505 DGNVYLTQTSYDLNKAAGVHYQFLYDYNVKPEVNDPKGNTSIEYADGKSVVFNINDKRN 564
DB 804 D-----IRNDFFIAFTVDSNDEVVDDVVEVNNKAKGMLA-EVTDKGMKDANDKMW 856
QY 565 NGFDEIGIEQHIIYINGKEYTSFNDIKQIIDKTLNIIKIVKDFARNVTVEFINKDTGEV 624
DB 857 VGLVSDVRKDTI-----TYKDADNKK-----KASIKSATV-YFDLYDDFGEA 898
QY 625 SELKPHRVTVTTIQNGKMSSTIVSSEDFILPVVKGLEKGYQFDGWEISGFEGKKGAGYV 684
DB 899 DG-----VNGDGVVMDISGDI-SGTXYD-----YV 923
QY 685 INLSKDTFTIKPVFKKIEE-----KKEENKEP-----TFDVSKKKONPQVNHNS 726
DB 924 LIVSDAKTVRK--DKLEDDAEAFKQEPSEKPDPTKWDALPSKVGKFTSAGPVKLYRA 981
QY 727 --QLNESHKEDLQREE--HSQKSDST-----KDVATV-----LQKNISSTSTNNPNK 773
DB 982 TVELNSKVKAEVDVAIEFYFNGKVEPSLLNFKDGVITIGYNTEDKVTSSKIKVTNKGK 1041

RESULT 14
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein.";
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wpa loci.";
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacry region.";
RL Microbiology 142:3113-3123(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brang A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Konigstein G., Krogh S., Kunano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina M., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L05634; AAA22883.1; -;
CC EMBL; D31856; BAA06656.1; -;
CC EMBL; D29885; BAA06260.1; -;
CC EMBL; D83026; BAA11683.1; -;
CC EMBL; Z99124; CAB15959.1; -;
CC FIR; S32920; S32920.
CC Subtilisin; BG10797; wapa.
CC InterPro; IPR003305; CBM_CentC.
CC InterPro; IPR006530; YD_
CC Pfam; PF02018; CBM_4_9_1;
CC TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
CC Cell wall; Repeat; Signal; Complete proteome.
KW SIGNAL 1 28 OR 32 (POTENTIAL).
FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 636 736 1-2.
FT REPEAT 769 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT REPEAT 1021 1040 2-1.
FT REPEAT 1042 1061 2-2.
FT REPEAT 1063 1082 2-3.
FT REPEAT 1083 1102 2-4.
FT REPEAT 1109 1128 2-5.
FT REPEAT 1129 1148 2-6.
FT REPEAT 1150 1169 2-7.
FT REPEAT 1174 1193 2-8.

FT REPEAT 1199 1218 2-9.
FT REPEAT 1219 1238 2-10.
FT REPEAT 1666 1685 2-11.
FT REPEAT 1667 1686 2-12.
FT REPEAT 1690 1709 2-13.
FT REPEAT 1711 1730 2-14.
FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24 (APPROXIMATE).
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2028 2047 2-27.
FT REPEAT 2051 2070 2-28.
FT REPEAT 2071 2090 2-29.
FT REPEAT 2093 2112 2-30.
FT REPEAT 2120 2139 2-31.
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;
Query Match 4.5%; Score 183; DB 1; Length 2334;
Best Local Similarity 20.9%; Pred. No. 1.2; Indels 320; Gaps 45;
Matches 195; Conservative 99; Mismatches 317;
QY 51 ISTIRDFENKDKLKKLKKF-----REYDDFTSETGKEMEYD----- 88
DB 1168 IQVIDDAEG--LKIITNTKYEGNVVDVDPNDVGTKATESYQYDKDGNVTSVKDAYGT 1225
QY 89 --YKY-----DDKONI--IAYDDGTDLEYETEKLDEIKSKI---YG-----V 123
DB 1226 ETEYKNKNDVTMKDKTEGNTVDIAY-DGLDAVSETDQSGKSSAAVYDKYGNQIOSSKD 1284
QY 124 LSPS-----KDGHEI-----LGIKNVSKNAKVYGNVKNYKSTIEI-----KAT 161
DB 1285 LSASNILKDGSEFAQKSGWNLTASKDRKISVIADKSGVLSGS--KALEVLSQSTSGT 1342
QY 162 KYDFHSKWTDFLYANI-----NDIVDGLAFAGDMRLFFKNDQKAEIKIRMEK 212
DB 1343 DHGYSSATQVLEPNTTYTLGKIKTDLAKSRAY---FNIDLRDKQK-----R 1389
QY 213 IKETSEYYPVSSYGNVIELGSDLSKNPDLNLTWESCKI-----YSDSEKQOYLK 265
DB 1390 IQWIRNEYSALA-----GKNDWTKRQITFTTPANAGKAVVYMEVDHKDKGKAWF 1441
QY 266 DNIILRKGYALKVTTVNPCKTDM-----LEGNGVYSKEDIKAIQKAPNLRALSET 316
DB 1442 DEVQLEKGEV--SSSNPVSQSSFTSATENNVNVSASVDSSEGF-----NDDVSLKAAR 1493
QY 317 TIYADSRNVDEGRSTQSVLSALDGFNIIRYQVFTFMNDKGEAIDKQGNLVTDSKLV 376
DB 1494 TSASQAGSV---TKQTVLQG-----SANDKPYLLTLTGMSKASSVKFT- 1534
QY 377 FKQDKKEY-----TGDDKFNVEAIKEDGSMFLFTDKPVLNLSMDK 415
DB 1535 ---DEKDYSLQANVTYADGSTGIYNAKFPSTQEWNRRAV-----VIPKTKPINKVDIS 1585
QY 416 NYFNPSKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYLGLHDINTDRDNI 475
DB 1586 ILFQKSATGVW-----FD-----DIRLEGSLLTKSTYDSNG 1618
QY 476 KLVNKGDIIMDMGK--DYKANGFPDPKVTMDG-----NVYLQTSYSLNA- 519
DB 1619 NYVTKEDELGYATSDYDTGKTSETDAKGEKTTYTYDQADQLTNWLTNLTSGTSLHSY 1678
QY 520 -----KAGVHVQFLYDNY-KPEVNIIDPKGNT-STIYADCKSVVFNINDKRNN 565
DB 1679 DKEGNEVSKIRAGADQTYKFYDVMGKLVKTTDPLGNVLAGEYDANSNLTKTISPNGNE 1738

QY 566 ---GPGETQEQHIYINGKEYTSFNDIKQIIDKTLNIIKIVKDFARNTTVKGFILNKOTG 622
Db 1739 VLSYDGTDRVRSKSYNGTEKIYFT-----YDKNGNETSVNK-EQNTTKKRTFDNK--- 1789
QY 623 EVSELKPHRVTVTIQNGKEMSSIVSEED-----FILPVYKG-----E 660
Db 1790 -----NRLTELDRGSGQTWTYPDSKDKLTFWSIHGDKQKTNQFTYKLDQMIEMKD 1842
QY 661 LEKGYQFDGWE-----ISGFEKGDAGY-VINLSKDTFI-----KPVFKIEEKEE 707
Db 1843 STSSYSFDYDENGWQTFITGNGGGTSFSDYERNLVSSHLHGDKNGGDLITSEYEDANG 1902
QY 708 NKPTPDVSKKQNPQVNHQNLNESHKEDLOREHSQ-----KSD 747
Db 1903 NRTTIN-SSASGKVQVEYKGLNQ-----LVKETHEDGTVEIYDGFGRNKTVTITKDG 1955
QY 748 STKDVATV-----LDKNISSKSTTNPNK 773
Db 1956 SSKTVNASFNMQ/LKVNDESISYDKNGNR 1986
RESULT 15
CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
COLLAGEN-CONTAINING SUBSTRATA.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81736; AAA20874.1; --
DR PDB; 1AMX; 24-JUN-98.
DR PDB; 1DZO; 27-SEP-00.
DR PDB; 1DZP; 27-SEP-00.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1154
FT PROPEP 1155 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
FT STRAND 174 179
FT TURN 182 183
FT TURN 185 186
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT TURN 239 246
FT HELIX 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT STRAND 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;
Query Match 4.5%; Score 179.5; DB 1; Length 1183;
Best Local Similarity 18.7%; Pred. No. 0.77;
Matches 172; Conservative 128; Mismatches 325; Indels 297; Gaps 42;
QY 22 KDDTTGVHHQENBESIEKSSFTIDRNISTIRDFENKDLK-----KLKKKREV 73
Db 304 KEENVKGSFNHTVHINAMAGIEGTVKGLVKL--QKDKTAPIANVKFLSKKDGSVV 361
QY 74 DDTSETGRMEYDYKYDDKG--NIIAYDDGTDLEYETE-----KLDEIKSIYGVLS 126
Db 362 KD-----NQKEIITDANGIANIKALPSGDYILKIEAPRPYTFDKKEYPFTMKDT 414
QY 127 SKDGHFEILGKISNVSKN-----AKVYGNKYSTEIKATKYDFHSTKMTFDLYANINDI 181
Db 415 DNOGYFTTIENAKAIEKTKDVSQAQWEGTQ-----KVQPTIY----- 452
QY 182 VDGLAFAGDMRLVFKNDQK-----KAEIKRMPEKIKETKSEYYPVSSYGNVIEL----- 232
Db 453 -----FKLYKDDNQNTTPVDKAEIKLEDDGTTKVTWNSLNPENDKNGKAIKYLKVE 503
QY 233 --GEGDLSKNKPDNLTKMESKIIYSDSEKQVLLKNDIILRKGYALKVTTYNPCKTDLML 290
Db 504 VNAOGE--DTTPEGYTKKENGVLVNTTEK-----PIETTSIS 538

```

QY 291 GNGVYSKEDIATQKPNLRLALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVF 350
Db 539 GEKVM-----DDKNDQDGRPEKSVNLLANGKVKTLDV 573
QY 351 TFKMNDKGRAIDKQGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVN 410
Db 574 TSETNWKYEF--XDLPKYDEGKKI-----EYT-----VTEDHVVDY--TTDIN 612
QY 411 LMDKNYFNPSKSNKIYVNR-----PEFYLRGKISDKGF-----NWE- 448
Db 613 GTTITNKYTPGETSATVKNWDDNNNQDGRPEIKVELYQDQKATGTAILNESNNWTH 672
QY 449 -----LRVNESVDNVLIIYG-DLHIDNTRDFNIKLVK-----DGDIMDWG 488
Db 673 TWGLDEKAKGOQVKYTVBELTKVGYTHVDNNDMGNLIVTNKYTPETTSISGEKV-WD 731
QY 489 MKDYKANGFPDKVT-----DMDGNVYLQTYGSDLNKAVG--VHYQFLYDNV 533
Db 732 DKNDQDGRPEKSVNLLADGEKVKTLDTVSETNWKYEFKDLPKYDEGKKIETVTVTEHDV 791
QY 534 KPEVNIDPKGNT-SIEYADGK---SVFNINDKRNNGPDG---EIQEOHIYINGK---- 581
Db 792 K-DYTTDINGTTITNKYTPGETSATVTKNWD--NNNQDGRPEIKVE-LYQDQKATGK 847
QY 582 -----EYTSFNDIKQIIDKTLNLIKIVVKDFARNTTVKEFILNKOTGE 623
Db 848 TAILNESNNWTHWTGLDEKAKGOQVKYTVBELTKVGYTHVDNNDMGNLIVTNKYTP 907
QY 624 VSEL-----KPHRVTVT-IONGKEMSSTIVSEEDFILPVYKGELEKGYQF 667
Db 908 TTSISGEKVWDDKNDQDGRPEKSVNLLANGKVKTLDTVSE-----TNWKYEFKDLPKY 963
QY 668 DGWEISGFEGKK-----DAGYVINLSKDTFIK-----PVFKKIEEKKEEENK 709
Db 964 D-----EGKKIETVTEHDHVVDYTTDINGTTITNKYTPGETSATVTKNWDNNNQDQK 1016
QY 710 -PT-----FDVSKKXDNQVNHSQLNESHKREDLQREHSQKSDSTKDVATV----- 756
Db 1017 RPEIKVELYQDQKATGTAILNESNNWTHWTGLDEKAKGOQVKYTVDELTKVNGYTH 1076
QY 757 LKNN-----ISSKSTNNPNK 773
Db 1077 VDNNDMGNLIVTNKYTPKXPNK 1098

```

Search completed: January 28, 2004, 13:00:44
 Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:57:43 ; Search time 20 Seconds
(without alignments)
3716.920 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGETAESKFNKLGNGKESG.....ATVLDKNISSKSTNNPNK 773
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	6.2	1850	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhoptry protein -
7	230	5.7	2401	2 T28676	rhoptry protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1840	2 A24594	probable major sur
11	218	5.4	1631	1 SAZQK1	major merozoite su
12	218	5.4	1639	2 S05603	lipoprotein (impor
13	217.5	5.4	1546	2 G90603	reticulocyte-bind
14	215.5	5.4	1252	2 B42771	DNA-directed RNA p
15	214	5.3	2339	2 A45597	membrane nucleas
16	212.5	5.3	1125	2 E90598	hypothetical prote
17	212.5	5.3	2166	2 G70163	serine/threonine-s
18	212.5	5.3	2485	1 H71621	protein with 5'-3'
19	212	5.3	1188	2 A71621	microbial collagen
20	211.5	5.3	1104	1 A36866	RESA-H3 antigen PF
21	209	5.2	1558	2 B71603	hypothetical prote
22	207.5	5.2	4688	2 E89824	hypothetical prote
23	207	5.1	1141	2 E89824	chromosome segrega
24	207	5.1	1191	2 B97116	hypothetical prote
25	207	5.1	1979	2 C71622	probable cell surf
26	206.5	5.1	1622	2 AE1717	mature-parasite-in
27	206	5.1	1526	2 A45605	RAD2 endonuclease
28	204.5	5.1	1516	2 E71619	hypothetical prote
29	204.5	5.1	2500	2 G71609	hypothetical prote

RESULT 1
F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95074
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hel-
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK7491.1; PID:g14972117; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

ALIGNMENTS

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 4e-167;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KLGETAESKFNKLGNGKESGSLKDDTGVHHHQHNEESIKESKSTIDRNI	TIRDFENK 60
DB	1334	KLGETAESKFNKLGNGKESGSLKDDTGVHHHQHNEESIKESKSTIDRNI	TIRDFENK 1393
QY	61	DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLVETEK	LDETKSKI 120
DB	1394	DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLVETEK	LDETKSKI 1453
QY	121	YGVLSPSKDGHEILGKISNVSNAKYVYNNYKSIIEIKATKYDFHSTMT	FDLYANIND 180
DB	1454	YGVLSPSKDGHEILGKISNVSNAKYVYNNYKSIIEIKATKYDFHSTMT	FDLYANIND 1513
QY	181	IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYPTVSSYGNVIE	IEGEGDLSKN 240
DB	1514	IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYPTVSSYGNVIE	IEGEGDLSKN 1573
QY	241	KPDNLTKMESKIVSDSKQQVLLKNDIILRGYALKVTTYNPGKTMLEGN	GVVSKEDI 300
DB	1574	KPDNLTKMESKIVSDSKQQVLLKNDIILRGYALKVTTYNPGKTMLEGN	GVVSKEDI 1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRY	QVFTFMNDKGEA 360
DB	1634	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRY	QVFTFMNDKGEA 1693
QY	361	IDKGNLVTDSSKLVFKDKKEYTGEDKFNVEAIKEDGSMLEFIDTKPVL	NSMDKNYFNP 420

Db 426 KIDQTEYVYKINKENYNDIILKNNLQKLEENKKIDQTEYVYKINKENY--NDIIE 483
Qy 270 LRKGYALKVTTNPG-----KTMLEGGVGYKEDIAKQKPNLRALSETTIYAD 321
Db 484 LKNNLQKLEENKNNKLTKLKNDIESNTELFNKLNSIDPKOKSREIAKLN--TEYEQ 541
Qy 322 SRN--VEDGRSQSVL-----MSALDGNRIYQVTFPMNDKGAIDKGNLVTDSK 373
Db 542 LRKDLLENKTNLMKLSNKLSSLE-----QLYDSKQNL-DGIDKIYNSLKEKN- 592
Qy 374 LVLFGKDDKEYGCEKFNVEAIKED--GSMFLFIDTPVNLMDKNYFN--PSKSNKIY 427
Db 593 -----DKIDEVFSNIEKFDIYVNIENKFIKINLSDSIINNOQFKEYINSKIDSNSLS 648
Qy 428 VRNPEFY-----LRKISDKGGFNWELRVNESVVDNYLIYGLIHIDNTRD 472
Db 649 TWFDDIFNAKNOIASITNNIENISNKID-----LNEFIISNEDSSKEL-LDEIRK 698
Qy 473 FNIKLVKQGDIMDMGMDKYKANGFPDKVTDMDGNV-YLOTGYSDLNAKA-----V 522
Db 699 YKQOQD-KIKDAMNTEVKSFE-NLQKIDISIKSNINELTNAYDIINTRKANDLDDKNNY 756
Qy 523 GVHYQFLYDNVXP-----EWNIDPKGNTSIEVAD-QKSVVFNINDKRNNGFDGEIQEH 576
Db 757 GSEFKMLYNASDLDTIOKNDEKVKQLNEYLEKQKOSIEINDIVNN-----805
Qy 577 YINGKEYTSFNIDKIQIDKTLNIIKIVVDFARNITVKEFILNKDTGEVSE--LKPHRYT 633
Db 806 FI--KELIKFNTE--TNKSLN-ELLTND--DINDKIFPKYKELNKISTNNLLKIYK--855
Qy 634 VTIONKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDAAGVINLSKDTF- 692
Db 856 NEIDNVNEXLSIVLENLQFINSFLSIEFNQ-----SITSH-----INFLNTLA 900
Qy 693 -IKPVFKLTKEEKEENKTFDFVSKKNDPQVNHQSOLNESHKEDLQREHSQKSDSTKD 751
Db 901 GINDVLNKLNLKIMADTTRRGDTNTR-----DEIKNOISSENIKSOKEKNE--KD 950
Qy 752 VTATV-----LDKNISSKSTNN 770
Db 951 LKKLISFNDKLNKYNISAGYTEYN 974

RESULT 4

T18444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A:Note: C0385c

Query Match 6.2%; Score 250; DB 2; Length 1650;

Best Local Similarity 20.8%; Pred. No. 0.0026;

Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

Qy 20 SLKKDTGTVHHQHEENEEKSSFTIDRNISTIRDFENKOLKLIKXKPRE---VDDP 76

Db 394 SLNENYNERKIYKEINKEYSNKEV-----HFKND-DSSIKKKNSSECLDRQ 443

Qy 77 TSETGKR--MEEDYKYDDKGNIIAY--DDGTDL-----YETKLDIKSKIYGVISPSKD 129

Db 444 KKKTYKTYIEQKRYNFRDNRNNNAVYIKDTHKKEKGYLNMIVQSEYKYTG--SNRKM 501

Qy 130 GHFEILGKIS-----NVSKNAKYVGNYYKSI--EIKATKYDFHSKTMTFDYANINDIV 182
Db 502 DMEIYNTQNTDNFINENLNKIYF--DDYEGYDPEKKKKLDDHIYTOQKEYKNNINDIL 560
Qy 183 DGLAFAGDMRLFKVNDOKKAEIKIRMPKIK-----ETKSEYPYVSSVGNVTEGEG-- 235
Db 561 K-----DHLNDKETKEKKKIEIEEEKKKNKIEIEEEKKKNKIEIEEEKK 614
Qy 236 ---DLSKNKPDNLTKMESGKIYSDSEKQOYLKDNII-----LR 271
Db 615 KKIIEIEEEKKKKI-EMEEENKIDDEKONTYANDKLIISHIDNVNCNIKIDALLDHIEKK 673
Qy 272 KGYALKVTTYNPGKTD-----MLEGGVYSKEDIAK-----IOKANPNL-- 310
Db 674 KTGHKEINLYKEIKNEYQKMLNDENSIMLEHEKYNKYNTHQVNNNLCDTKMLQKENKILT 733
Qy 311 RALSETTIYADSRNVEDGRSTOSVLSALDGNRIYQVTFPMNDKGAIDKGNLVT 370
Db 734 NDKKTFLLSKSKNI-----TSNVLSSKIPG-----TLSTKCLNATIKTIKKD--VTD 779
Qy 371 SSKLVLF-----KODKE-----YTGEDKE-NVEA-----IKEDGSMFLFIDTK 407
Db 780 NEKKYVVDHHRKDNIIKKNKEFINIYKGNANVANVEIGSEVCNKNVNVKDDNMVENVK 839
Qy 408 P-----VNLSDMKNYF-----NPSKSNKIYVRNPEF 433
Db 840 QGDDNNMVMVENKQDNNVIVKNIERSESSFMFTHRKNITSTSTDTCAKNEQIKYPHL 899
Qy 434 YLRGKISDKGGFNWELRVNESVVDNY-----LYDGLHIDN-----469
Db 900 Y-SNRKEDKOKKSIIFLKNINENIKKNYKDKKISTLEKKVFKDKNVITNDDDKHTSSK 958
Qy 470 -TRDFNI-----KLN-----478
Db 959 INDDFNITIDKQKGLNNPVDLNRIKIKETKLEKDKSHMSKIQNNLMKKTNPTNNK 1018
Qy 479 -VKGDGIMDMGMDKYKANGFPDKVTDMDGNVYL-----QTG 513
Db 1019 GISSTSISSSTKFNKFCGIEKNKNSLNLKYTCIRKNMNVDSIKLNDKADLYKDKKTS 1078
Qy 514 YSDLNKAVGHYQFLYDNVNVKPEVIDPKGNTSIEYADKSVVFN-----INDKRNNGFDG 569
Db 1079 FNDINRAAKGMNFK-----KRDV-----PNKNMNVDTNKGKR-VFNPVTLLNNYRNN-- 1124
Qy 570 EIQEQHIYINGKEYTSFNIDKIQIDKTLNIIKIVVDFARNITVKEFILNKDTGEVSELKP 629
Db 1125 -----YIRSNK-----NNVK-----NGRW-----GIKKILVLKE--KQKSLHP 1156
Qy 630 HRVTVTIONGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDAAGVINLSK 699
Db 1157 EGVEADKKLNSYNNDKYLIEKDGFKDIIINEEMEK-----YKNNK-MKYKI--K 1201
Qy 690 DTFIKPVPFKKIEEKEENKPTFDVSKKNDPQVNHQSOLNESHKEDLQRE--HSQKS 746
Db 1202 SNSIPIIKKIERKSNNDN---DNKNDNINSNNNINSNDKCLFLSKERDRVHLKNN 1258
Qy 747 DSTKDVATVLDKNNISSKSTNNPNK 773
Db 1259 NIIVNTMMFRKQSNSCDNNTTSLKNN 1285

RESULT 5

T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CABL1104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 5.8%; Score 235.5; DB 2; Length 3724;
Best Local Similarity 18.7%; Pred. No. 0.032;
Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

QY 5 IAESKFKNLGNGEGS-LKK-----DTTGVHEHGOEESIK-----BKSSFT 46
DB 249 IPKNLFENLLNKHNDYLRNILLMDVNDINPLEHPDEQNSLNKCLTGNKKYKM 308
QY 47 I-----DRNISTIRDFENKDLKLIKKKPREVDFTSETGKRMEEDYKYDDKG 95
DB 309 IPKGTBYQDKEKESILTINQNDKKYKKK-KKKYSELQDSNISSNNTLTSKRYTYT--- 363
QY 96 NIIAYDDGTDLEYETEKLEIKSIYGVLSPSKDGHPFELGKISVSKNAKVYGNKYKS 155
DB 364 -----CGMDKETKETDEQNK-----RNTVASICNVCVNETNEKORKKTKKK 407
QY 156 IEIKATKYDFHSKTMFTFDLYANIDVDGLAFAGDMR--LFVKDNDQKKAETKIRMP-EK 212
DB 408 ---SVQKDLVEENALLD---NL-IIIDGINFDDVTKECKIIDDNNNENIDNNIYVENNK 460
QY 213 IKETKSEYPVSSYG-NVI-----ELGBGDLGSKNPDNLTKMESGK 252
DB 461 LKDKQS-YDLFSSEGNLILGVNEGEEFNEEFENIEKELQERKDEKKNK-----T 512
QY 253 IYSDSEKQOYLKDNILLRKGVALKVTYTPGKTMLEGNVYSGEDIAKIQAN-PNLR 311
DB 513 IYNNNEEQTDLNRNT-----NKIESINNNNDN---NNNINNKKEFNKIRTEHILNKE 562
QY 312 ALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMDKGEAIDKDGMLVTS 371
DB 563 SISKHKIGSPSRDK-----IKKLYTNKNEDSTFELKKELEIITNN 604
QY 372 SKLVFGKDKYTGEDKFNVEAIKEDGSMPLFDITKPVNLSMDKNVFNPSKN----- 424
DB 605 KVVVY--EEDIIGSNEDDEVIHLKE-----NLKEDANEYNDKENKKNKKE 650
QY 425 ----KIYVRN-----PEFYLRGKISDKGFNWLRYN---ESVVDNYLYGLDHLIDTDD 472
DB 651 ILKSKNYLENKRTLEELKRGK---NNIFKDKETNSLGEVINEIQINEENKINDIQD 707
QY 473 FNIK-----LNVKD---GDIMDMGMDKYKANGPDKVTDMD----- 505
DB 708 GNISKOKIIQSSRTWDTFNKIDISLNDLLEKREKKKSKQHFDNLVLKADKNEISENINK 767
QY 506 -----GNVYLQT-----GYSDLNKAVGVHY----- 526
DB 768 ICDNNINNIYDESINNIYDESINNIYDESINNIYDENINNIYDENINNIYDE 827
QY 527 -----QFLYDNVPEVNI-----PKNTSIE-----Y 549
DB 828 NINNIYDEGINKICDDNILENKNIKTNDIVQVEENNESIEKNELMISLNKOINNTYNNP 887
QY 550 ADGKSVFN-----INDKRNNGFDGEIOE----- 573
DB 888 KENVDFPNKIKRESLLKDKNKKDNNDDEYIMDNFYENDFINHMKBITNKELDPLRI 947
QY 574 -----QHIYINGKEYTS---FNDIKQIIDKTLNIIKVVKOPARNVTYKFIPLNKDTGE 623
DB 948 NTQNEFIENLDIKKKYTYNDHPFNADAKMFE-MN-KILNKKMKKKEQEFKTDFTFGS 1005
QY 624 VSELK-----PHRVTVTIQNGKMSSTIVSE 649
DB 1006 LQSHKIKKYNKGEEKDKNNEEKNILYDENQVSVLYSDHKIEQDIOIHISQTNICDE 1065
QY 650 EDFILPVYGELEKGYQFDGWEISGPEGKADAGVNLNLSKDTIKPVFKKIEBKEEENK 709

DB 1066 NN-IEQINEENS KGVRIISGDTM-----ENKND-----MENKNDMEKK 1102
QY 710 PTFDYVSKKNDPOVNH-SQLNESHKEDQLREHSHQSDSTKDVATVLDKNINSSKSTT 768
DB 1103 N--DMEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDE 1153
QY 769 NNPKN 773
DB 1154 ENENK 1158

RESULT 6
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
A:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: Z20508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
A:Cross-references: EMBL:L27838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 5.8%; Score 231.5; DB 2; Length 2269;
Best Local Similarity 19.7%; Pred. No. 0.025;
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;

QY 37 ESIEKSKSFTIDRNTSTIRDFENK-DLKKLIKKKPREVDFTSETGKRMEY----- 87
DB 109 EGLKLELNKKIILAKIEYVKNVTVELKEIEKNAYIDELANOSPKYKVTGIENKNTIY 168
QY 88 -----DYKYDDKGNIIAYDDGTDLEYETEKLEIKSIYGVLSPSKDGHF 132
DB 169 NTIKSYFOIYEGDITDFYNELSSIVKEDPIDDIEDKT-KLENLRSKINDVYDKIQKMEI 227
QY 133 E-ILGKISNVSKNAKVYGNKYKSIKATKYDFHSKTMFTFDLYANIDVDGLAFAGDM 191
DB 228 ETVKSHLNINETNNKL-----PNTILEIKYIYDEISK-----ELNKMLEDKKNK---EKEL 276
QY 192 RLUFVKNDQKK--AEIKIRMP-----KIKETKSEYPPYVSSGVNVELGE 234
DB 277 SNKISDYDKRQLSEYKSKMLEIRNHNVSQTNVDNTEKEEAKQNYDKSNEHMTTIPTNE 336
QY 235 GDLK-----NKDNLTKMES-----GKIYSD-----SEKQOYL-LKONI---ILRKYAL 276
DB 337 DEISKIIISVTKMDEILSKVNTYIDFNKYYKETVNSEHSTQFTELTDKIAEVSDEKELKK 396
QY 277 KVTYNPGKTMLEGNVYVSK--DIAKIQANPNLRALSET-----TIYAD-- 321
DB 397 CQSFNDNKSILNETKNSIEKQYQNTLKKVDEYIKVCKSTKESITKFSKQTLKMDL 456
QY 322 SRNVEDGRSTQSVLMSALDGFNIIRYQVET-----FKMN-----DKGEAI--- 361
DB 457 NQNIKTVKETSIDKSYIEKFE---QILTGRQTKLENKFTBFSLNHEANNNELIKYFS 512
QY 362 KDKGHLVTDSSKLVLFPGKDDKEYTGEDKENVEAIKEDGSMPLFDITKPVNLSMDKNVFNPS 421
DB 513 DLKANIGINEENNLVYNQFTEKEKTFND-----IKEKNIHINEEISKIEIKHASIYNIS 566
QY 422 KSKNIYVRNPEFYLRGKISDKGFNWLNVSVVDNYLYIGDLHIDNTRDFNKLNVK-- 480

A:Reference number: A24594; MUID:86014355; PMID:2995820

A:Accession: A24594

A:Molecule type: DNA

A:Residues: 1-1640 <HOL>

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640;
Best Local Similarity 18.1%; Pred. No. 0.056;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

```
QY 23 KDTTG-VEHHQNEESIK-----EKSFTIDRNISTIRDPENKDLKLIKKEPREVDD 75
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 KDNVGMEDYIKNNKKTITENINELIEESKTTIDKNNATKEEB-----KKGLYQAY 283
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 FTSETGRMEEDYKYDDKGNIIAYDDGTDLEYETEKLEIKSIKIYGLVSPSKDGHFEIL 135
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 DLSIYNKQLEE-----AHNLISV-----LE---KRIDLK-----KNENIKELL 319
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 GKISNVKNKAKVYVGNKYKSIIEIKATKYDFH-----SKTMTFDLYANINDIVDGLAFA 188
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 DKINEIKNPPANGSNGTPTNLLDKNKKEIIEHEKEIETIKFNIDSLFTDPL----- 373
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 GDMRLPVKNDOKKAIKIRMEPEKIKETKSEYP-----YVSSYGNV-----IEIGE-GDLSK 239
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 -ELEYLREKN-KNIDISAKVETKESTEPNPGVTYPLSYNDINNALNELNSFGDLI- 430
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 NKPDLNLTWMSGKIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYNPGKTDML 289
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 -NPFDTYKPSKNITDNERKFFINEIKIEKIEKKIESDKKSYEDRSKSLN-----DIT 485
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 EG-----NGVYSKE-----DIAKIQA-----NPNRALSETTIYADSRNVEDGRSTQSV 334
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 KEVEKLNLNIEYDSKFNNNIDLTNFEKMGKRYSYKVEKLTHPTNFTASYENSKHNLEKLT 545
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 LMSALDGF--NII-----RYQVFTFKNDGGEALDKGNLVTDSKLVLFKGDKEYT 385
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 ALKYMEDYSLRNIVVEKLYKYLKSLKSIENIETLVE--NIKKDEQLF-----EKKIT 598
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 GEDKFNVEAIKEDGSMFLFTDKPVLNLSMDKNVFNPSKSNKIYVRNPEFYLGRKISDKG 445
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 KDNKPEDEKILEVSDIVKQVKVLL---MNKIDELKKTQLILKNVE--LKNHIVPNSY 653
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 NMLRVNESVVDNLYLYGDLHIDNTRDFNIKL-----NVK----- 480
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 654 KQENKQE---PYLLVLKKEIDKLKVPKVESLINEEKNKIKTQGGSDNSERPTEGEI 709
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 481 -----DGDIMDWGMKDYK-----ANGFPDKV 501
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 710 TGOATTKPGQAGSALGDSVQAQEQKQAPPPVPVPEAKAQVPTPPAPVNNKNTNV 769
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 502 TDMGDNVYLQGTGYSDLNA-----KAVGVHYQFLYDNVKNPEVNIIDPKGNISIEYADG 556
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 770 SKLD---YLEKLYQFLNTSYI CHKYLVSHTMNEKILKQYKITKEESKLSGCCPLDILL 826
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 557 FNIN-----DKRNNPGD---EIQEQHIYINGKEYTSFNIDIKQIIDKTLIKIYV 603
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 827 FNIQNNIPWYSMFDSLNNLSQLFWEIYEKEMVCNLYKLKNDKIKNLLEAKKVSISV 886
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 KDFARTTVKEFILNKDTEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYGELEK 663
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 887 KTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSLKFENILS-----LGNKN 937
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 664 GYQFDGWEISGEGK-----KQAGYVNLISKDTFFIKPVFKIIEKEEENKPTF--D 713
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 938 IYQ-----ELIQKSSNFYEKILKQSDTFTYNESFTNFKSKADDINSLNDESKRKLLED 993
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 714 VSKKDNPOVNSQLNESHKEDLQREHSQKSDSKDVTATVLDKNNISSK-STTNPN 771
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 994 INKLKKTQLQSPDLNKNYKYLKLERLFDKKTGVKQKQIKLTLILKEQLESKLSLNNP 1052
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 11

SAZQKI

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (s

C:Species: Plasmodium falciparum

C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000

C:Accession: A25120

R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.;

EMBO J. 4, 3823-3829, 1985

A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium fal

A:Reference number: A91030; MUID:86136024; PMID:3004972

A:Accession: A25120

A:Molecule type: DNA

A:Residues: 1-1631 <MAC>

C:Comment: The merozoite stages of different strains have strain-specific surface antig

C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The m

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prot

F:1-19/Domains: signal sequence; status predicted <SIG>

F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>

F:67-84/Region: 3-residue repeats (S-G-T/P)

F:1614-1631/Domains: membrane anchor #status predicted <MBN>

F:97,259,755,759,835,911,955,1049,1156,1436,1563/Binding site: carbohydrate (Aen)

Query Match 5.4%; Score 218; DB 1; Length 1631;

Best Local Similarity 18.1%; Pred. No. 0.062;

Matches 164; Conservative 162; Mismatches 339; Indels 240; Gaps 37;

QY 23 KDTTG-VEHHQNEESIK-----EKSFTIDRNISTIRDPENKOLKK-----LIKKK 69

||| :

Db 223 KDNVGMEDYIKNNKKTITENINELIEESKTTIDKNNATKEEBKKLYAQVDFLYNQ 282

||| :

QY 70 FREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEYETEKLEIKSIKIYGLVSPSKD 129

||| :

Db 283 LEEAHNLISVLEKRID-----TLKKNENIK----- 307

||| :

QY 130 GHFEILGKISNVKNKAKVYVGNKYKSIIEIKATKYDFH-----SKTMTFDLYANINDIV 182

||| :

Db 308 ---ELLDKINEIKNPPANGSNGTPTNLLDKNKKEIIEHEKEIETIKFNIDSLFTDPL 364

||| :

QY 183 DGLAFAGDMRLPVKNDOKKAIKIRMEPEKIKETKSEYP-----YVSSYGNV-----IEIGE 234

||| :

Db 365 -----ELEYLREKN-KNIDISAKVETKESTEPNPGVTYPLSYNDINNALNELNS 416

||| :

QY 235 -GDLSKNKPDNLTKWMSGKIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYNP 283

||| :

Db 417 FGDLLI--NPFDTYKPSKNITDNERKFFINEIKIEKIEKKIESDKKSYEDRSKSLN- 473

||| :

QY 284 GXTDMLEG-----NGVYSKE-----DIAKIQA-----NPNRALSETTIYADSRNVEDG 328

||| :

Db 474 ---DITKEYEKLNLNIEYDSKFNNNIDLTNFEKMGKRYSYKVEKLTHHTNFTASYENSKGN 530

||| :

QY 329 RSTQSVLMSALDGF--NII-----RYQVFTFKNDGGEALDKGNLVTDSKLVLFPGK 379

||| :

Db 531 LEKLTALKYKYMEDYSLRNIVVEKELKYKYLKSLKSIENIETLVE--NIKKDEQLF----- 584

||| :

QY 380 DDKEYTGEDKFNVEAIKEDGSMFLFTDKPVLNLSMDKNVFNPSKSNKIYVRNPEFYLGRKI 439

||| :

Db 585 -EKKITKDNKPEDEKILEVSDIVKQVKVLL---MNKIDELKKTQLILKNVE--LKNH 638

||| :

QY 440 SDKGFFNMLRVNESVVDNLYLYGDLHIDNTRDFNIKL-----NVK----- 480

||| :

Db 639 HVPNSYKQENKQE-----PYLLVLKKEIDKLKVPKVESLINEEKNKIKTEGSDNSRP 694

||| :

QY 481 -----DGDIMDWGMKDYK-----AN 495

||| :

Db 695 STEGKITQATTKPGQAGSALGDSVQAQEQKQAPPPVPVPEAKAQVPTPPAPVN 754

||| :

QY 496 GPPDKVTMDGNNVYLQGTGYSDLNA-----KAVGVHYQFLYDNVKNPEVNIIDPKGNISIEYA 550

||| :

Db 755 NKTENVSKLD---YLEKLYQFLNTSYI CHKYLVSHTMNEKILKQYKITKEESKLSLSSC 811

||| :

QY 551 DGKSVVFNIN-----DKRNNPGD---EIQEQHIYINGKEYTSFNIDIKQIIDKTL 597

||| :

Db 812 DPLDLLFNINQNNIPWYSMFDSLNNLSQLFWEIYEKEMVCNLYKLKNDKIKNLLEAK 871

||| :


```
QY 175 YANINDIVGLAPAGDMRLFVNDQKAEIKIRMEPEKIKETKSEYPPYVSSYGNVIELGE 234
Db 268 Y--IEKIDOLFMAST--IKDTYO-----TNKIFLTNDY----- 298
QY 235 GDLKXNKPDLTME-----SGKYS-----DSEKQOYLLKNDLILRGYAL 276
Db 299 --IKGNESNLKDKWFSNNFENEIYKYFLEKLDALVEIQALHKQYLEKNEIILKTGEIVD 356
QY 277 KYTTYNPGKTDLENGG--VYSKEDIKI-----QKANPNLRALS 314
Db 357 KVAAPTKSRELKSKAGLLFSKQDSAKITQLINHILSRYPPEAPENLELSKAKMGLINEL 416
QY 315 ETIIVADSRNVEDGRST--QSVLMSALDGNFIIRYQVFTFMNDKGEAIDKGNLVTDSK 373
Db 417 KPEIYVEKEGVIAYPTLODAISNAQDGQKIFL-----NKNLKLK--SIVVDKN-- 464
QY 374 LVLFGKDKKEYTGED--KFNVEAILEDGSMFLFDTKPVNLSMDKNYFNS----- 421
Db 465 ITIFAKSNVTIIRKOSKSFTHFIYQKGAITFEIAEPSSQSNLNLGLGTSFKDESSLVKI 524
QY 422 -KSNKIYVRNPFYLRGKISDRGGFNWELRVNESVVDNYLYYGDLDHIDNTRDN----- 474
Db 525 EKNAKLVAKGTAFINSKFSKYG-----SVFEN--YGSVIEGAKIWNVSESG 572
QY 475 -----IKLVKGDIDMDGMDKYKANGFPDK-VTDMGDNVYLOTGYSDLNKAVGVH 525
Db 573 GIIRNVGSSSLTFKNGEI-----RDNISGT--DKGIYISQGNIAISGSDGNKS----- 620
QY 526 YQFLYDYNKPEVNIIDPKNGTSEYADGKSVVNINDKR-----NNGFDGILQCHYIING- 580
Db 621 FRSSLINLE-KTNINPNSGSIYVNNASVKSILFEIDNSKIOISNNALINPFGSAIFLKN 679
QY 581 -----KEYTSFNIDIKQIIDKTILNIIKIV----- 602
Db 680 STMHLAGSLKIKKEASEQIEVLVDLPQAKLISPKNIISLDNYQLSSAIFKIFSVCN 739
QY 603 VKDPAR-----NTTVKEFILNKDT-----GEVSELKPHRVTVTIQNGKEMSTIVSEE-DF 652
Db 740 INDFKHVPLVMTNKEKFFKLPWDTKLFVNFYKTLKQNDHLLQSGDGFESTEKIIKDELDF 799
QY 653 -----ILPVYKLEKGYOP-DGWEISGEGKK-----DAGYVINLSK 689
Db 800 YFRPTAAVKKLILTLQVTRTIPYKHFWEAFYVNFPLQKWETINELIRLDFPYLFDIAY 859
QY 690 DTFIKPVPFKIBEXKEEENKPTFDVSKKDNPOVNSQLNSHRKEDLQR 739
Db 860 PEFV-----ENGKMLKPEY-IHTNVNVPVLEH-----FRNEDVAR 894

RESULT 14
B42771
reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
C:Species: Plasmodium vivax
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997
C:Accession: B42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338; PMID:1617731
A:Accession: B42771
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1252 <GAL>
A:Cross-references: GB:M88098; NID:gl60627; PID:gl60628
A:Experimental source: strain Belem, merozoites
C:Genetics:
A:Gene: RBP2
```

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
Best Local Similarity 20.9%; Pred. No. 0.056;
Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;
QY 2 LGIEAESKPK--NLNGKESGLKDDTTGVGHHHQENESIKESST-IDRNI----- 51

RESULT 15

A45597

DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000

```
Db 266 LSEIKYKDKCTEISNKR-----KDKIEFLKFKPNEESNKNVINEINIRNSEY 322
QY 52 -----STIRDFENKDLKLIK--K--FREVDDFTS 78
Db 323 LKDIEDAQKASTKVLPFKHETTTISNIFKSEELIGVETKSQKKINKAEDIMKEIRHNS 382
QY 79 ETGKMEYDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKIYGVLSPSKDGHEFLGKI 138
Db 383 ETIQTKVGQFQENLKNLNEPHNYDNAED--ELANDKSTNAKVLJETNLESVKHN--LSEI 437
QY 139 SNVSKNAKYVYGNVYKSI--EIKATKYDPHSHKTM-----TFDLYAN-INDIVDGLAFAGD 190
Db 438 TNIKOGGEKIY--SKAKDIMQKIKATSENTAEKTELEKVKDDQSNVYVNLQIIT--E 490
QY 191 MRLPVKDN-----OKKAEIKIRMEPIKE-----TK-----S 218
Db 491 RNLIVTEKRNRLNGIDSTITNIEGALKESGNYEIGFLEKLEIGKRNKRLKVDITKKSINS 550
QY 219 EYPYVSSYGNVIELGEGDLSKKNPNLTKMESKGIYSSEKQOYLLKNDIILKRGYALKV 278
Db 551 TVGNFSSLFNFDLNLQYDFNKNINDYENKM--GEIYNEPEGSLNKISENL--RNASENT 605
QY 279 TTYNPGKTDMLEGNGVYSKEDI---AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
Db 606 SDYNSAKTLREA---OKEKYNLLNKEEANKYLR-----DVKKVESFR----- 646
QY 336 MSALDGFNIIRYQVFTFMNDKGEAIDKGNLVLTDSSKLVLFKGDKKEYTGEDKFNVEAI 395
Db 647 -----FIFNMK-----ESLDKINEMIKKEQLTV-----NEGHGNVKQVLVENI 683
QY 396 KEDGSMFLFDTKPVNLS-----MDKN-----YFNPSKNKIYVRNPFYLRGKI 439
Db 684 KE-----LVDEN--NLSDLKQATCKNEEIQKITHSTLKNKAKTILGHVDTSAKYVGIKI 736
QY 440 SDKGGFNWELRVNESVVD-----NYLIYGDLDHIDNTRDNKILNKVD--- 481
Db 737 TP-----ELALTTELLGDAKLTAQELKPFESKNVNVLETENMSKNTNLDVHKNIQDAYK 790
QY 482 -----GDIMDMGMDKYKANGFPDKVTDMDGNVYL-----QTGYSDLNKAKAVGVH 525
Db 791 VALETLAHSDEIDTKQD-----SSKLIEMGNQIYLVKVLINQYKNKISSIKSEEAUS 844
QY 526 YQFLYDYNK-----PEVNIIDPKNGTSEYADGKSVVFN-----NDKRNNGFDGEBIQBO 574
Db 845 VKI--GNVSKHSELSKITCSKSYDNIILALEKQELQNLRSFTQEKNTNTNSDSKLE-- 900
QY 575 HIYINGKEYTSFNDIKQIIDKTILNIIK-----VVKD----- 605
Db 901 -----KIKTDFESLKNAL-KTLEGEVNALKASDNHEHVQSKSBFVNPALSEIEKEETD 953
QY 606 -PARNTTVKEFTLNKDTGEVSELKPHRVTVT--IQNGKEMSTIVSEEDFILPVYKGELE 662
Db 954 IDSLNTALDELLKKKRTCEVSRYKLIKOTVTKEISDDTELINTIEKN-----VKAYLAYIK 1009
QY 663 KYQFDGWEISGEGKDGAGYVINLSKDTFFIKPV-----FKKIEEKKKEENKPTFD-- 713
Db 1010 KNYE-----DTVQD---VLTNEHFNTKQVSNHEPTNFDKNGSKSEELTKAVTDSK 1057
QY 714 --VSKKK-----DNPQVN-----HSQLNESHK--KEDLQREHSHQ 744
Db 1058 TTIISKLGVIIEVNTENTMTTIESAKIEALYNELKNKNTSLNEIYQTSNEVKLQEMKS 1117
QY 745 KSDSTKDVTA--TVLD---KNNISKSSTNN 770
Db 1118 NADKIIDVSKIFNTVLTQKSNIVTNQHSINN 1149
```

C:Accession: A45597
R.Li, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 46, 229-239, 1991
A:Title: Characterization of the gene encoding the largest subunit of Plasmodium falciparum
A:Reference number: A45597; MUID:92018020; PMID:1656254
A:Accession: A45597
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2339 <L11>
A>Note: sequence extracted from NCBI backbone (NCBIN:61099, NCBIPI:61101)
C:Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
C:Keywords: nucleotidyltransferase; transcription

Query Match 5.3%; Score 214; DB 2; Length 2339;
Best Local Similarity 21.6%; Pred. No. 0.15;
Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

QY 24 DTTGVHHQEE--NEESIKESGFTIDRNISTIRDP---ENKDLKKLIKKEPREVDDFTS 78
DB 1292 DNTYVEQIEKELSKNKKTEKQSPK-----GTIRDHEDSEBQMNFITKKAQ---FFI 1342
QY 79 ETGKMEBYDYKDKGNIIAYDDGTDLEYTEKLDDEIKSIY-----G 122
DB 1343 EK-KGKMEHCNDIEYNTQYD---NIQYNNISCNVYKSNLENTHTHQNNDLSFIKNN 1398
QY 123 VLSPSKDGH-----FELGKISNVKNAKYVYNNYKSIKATKYDFHKTMTFPL 174
DB 1399 VILPPKEYHSIFHFVNDYRVNVEIKMLDKKIFLNNSEKN--VVOVKYRMRMSKNLKKI 1456
QY 175 YANINDIVDG-----LAFAGDMRLFVKNDQ--KKAETKIRMPK-----IKETSE 219
DB 1457 EI-INNIYRNEKKLNKRWTKMDNDNYWSSDDSIIAKKIKIKNKEKKYHPKEKEN 1515
QY 220 YPVVSYGNVIELGEGDLKKNKPDNLTK-MESGKIYSDSEKQOYLKDNIIILKGYALKV 278
DB 1516 FDR-NNYKMITDNNNDNNNNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1569
QY 279 TTYNPKTMDL--EGNGVYSKEDIKIQANPL-----RALSETTIYADS-- 322
DB 1570 TNYN---TNIYNDNCNGIYEKETNNNELTNSNMCNDKNDFDEFFNNINENDLLYDNKY 1626
QY 323 -----RNVEDGRSTQVLSALDGFNIIRYQVF---TFKMDKGEAIDKGNLVTDSKL 374
DB 1627 YRQIFKNVIGFVSFVYVESYKQHYILFPYEIIKWTSPLELYTEIPTNIFLHTKLSK- 1685
QY 375 VLFGKD---DKYTGDBKFNVEAIEDGSMFLFDTKPVNLSMDKNYFNPSKKNKIYVRNP 431
DB 1686 ---KEKPTHQKNTGKMKIYIEIKK---WLFI--KAINI---YKYVSPKKSIEL----- 1728
QY 432 EFYLRGKISDKGFGNHELAVNESVVDNLYIGDLHIDNTRD-----FNI-----KLVKD 481
DB 1729 -----IKKQYFNIIK-NYDISHRYIIH-DYSPINLQKLYLFFNFIYKYFYISTP 1779
QY 482 GOIMDMGMDYKANGFPDKVTMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNIDP 541
DB 1780 GD---AVGSIASAQSGEPTQMTLTFHFAGVASMNV-TLGV-----PRIKEIINASN 1828
QY 542 KGNITSYADGKSVFVNINDKRNNGF-----DGEIQOHYIYNGKEYTS-----F 586
DB 1829 SIQTPI-----LNIPLEVDNPNFALMMKSKLEKTTIRDCMYIK-EDYTSRGVFLSVKF 1882
QY 587 ND-----IKQIIDKTLNI-----KI---VVKOPARNTVK--EFILNKDT 621
DB 1883 NEELIQKLFNINAYNIKQIILKQSHINKIKINKIHINVINKYKHLHISLKNDEFIF---- 1938
QY 622 GEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWEISGFEKG--K 679
DB 1939 FQMESLKGLLDLLIYGDKDKICEIKKEDI-----EVTNDEDEICD 1980
QY 680 DAGYVNLNKDTFIKPVFKIEKK---BEENKPTFDVSKKDNPNQVNHSQLNESHKED 736
DB 1981 DMDEYYNVSQGT-----ELYERKCNKEENKNAIRVKKEID-----DN 2019

QY 737 IQREB-----HSQKSDSTKDVATATVLDKNNISSKSTNN 770
DB 2020 LEKEENIIYVSEKDSVQNLKSEKKDIN-----DDNNNDNNNNN 2060
Search completed: January 28, 2004, 13:02:11
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:56:43 ; Search time 42 Seconds
(without alignments)
4749.397 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAERKFNLGNGKEGS.....ATVLKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2119	2 Q9AHT5	Q9Aht5 streptococc
2	4026	100.0	2140	16 Q37RY6	Q37ry6 streptococc
3	3793.5	94.2	2144	2 Q9S4M8	Q9S4m8 streptococc
4	3709.5	92.1	2144	16 Q8DQP7	Q8dqp7 streptococc
5	265.5	6.6	1642	5 Q8IB84	Q8ib84 plasmodium
6	264	6.6	3504	5 Q8IL45	Q8il45 plasmodium
7	255	6.3	1777	5 Q8I3P4	Q8i3p4 plasmodium
8	253.5	6.3	1127	12 Q9YVT6	Q9yvt6 melanoplus
9	252.5	6.3	2227	5 Q8II21	Q8ii21 plasmodium
10	251.5	6.2	2849	5 Q8IH4	Q8ih4 plasmodium
11	251.5	6.2	3381	5 Q8I2V4	Q8i2v4 plasmodium
12	251.5	6.2	3519	5 Q8IE65	Q8ie65 plasmodium
13	250	6.2	1474	5 Q8IIU2	Q8iiu2 plasmodium
14	250	6.2	1650	5 Q7J328	Q7j328 plasmodium
15	247.5	6.1	5767	5 Q8I525	Q8i525 plasmodium
16	246.5	6.1	2033	5 Q8IM18	Q8im18 plasmodium

17	245	6.1	2269	5 Q8ILA2	Q8ila2 plasmodium
18	242.5	6.0	3317	16 Q8EWP8	Q8ewp8 mycoplasma
19	241.5	6.0	1033	5 Q8IBB8	Q8ibb8 plasmodium
20	240.5	6.0	4433	5 Q8IJ15	Q8ij15 plasmodium
21	240	6.0	1455	5 Q8IKG8	Q8ikg8 plasmodium
22	240	6.0	10061	5 Q8I3Z1	Q8i3z1 plasmodium
23	239.5	5.9	2569	5 Q8IBG8	Q8ibg8 plasmodium
24	238	5.9	1081	16 Q8XIL2	Q8xil2 clostridium
25	238	5.9	3322	5 Q8IKL0	Q8ikl0 plasmodium
26	238	5.9	3628	5 Q968Y0	Q968y0 plasmodium
27	238	5.9	3704	5 Q8IKY8	Q8iky8 plasmodium
28	237.5	5.9	3535	5 Q8IC29	Q8ic29 plasmodium
29	237	5.9	2511	5 Q8IL44	Q8il44 plasmodium
30	236.5	5.9	1114	5 Q97242	Q97242 plasmodium
31	236.5	5.9	1738	5 Q8IAL5	Q8ial5 plasmodium
32	236.5	5.9	4044	5 Q8IDD4	Q8idd4 plasmodium
33	236	5.9	2228	5 Q8I1W5	Q8i1w5 plasmodium
34	236	5.9	3620	5 Q968T6	Q968t6 plasmodium
35	235.5	5.8	1692	5 Q8IJD6	Q8ijd6 plasmodium
36	235.5	5.8	3724	5 Q7J320	Q7j320 plasmodium
37	235	5.8	1936	5 Q8ISA6	Q8isa6 plasmodium
38	235	5.8	2184	5 Q8ILO6	Q8ilo6 plasmodium
39	234	5.8	2082	5 Q8I2R1	Q8i2r1 plasmodium
40	234	5.8	2359	5 Q8I297	Q8i297 plasmodium
41	234	5.8	2760	5 Q8ISY2	Q8isy2 plasmodium
42	233.5	5.8	2752	5 Q9BJY0	Q9bjy0 plasmodium
43	233.5	5.8	3610	5 Q968T7	Q968t7 plasmodium
44	232.5	5.8	7170	5 Q8IL30	Q8il30 plasmodium
45	232	5.8	2361	5 Q8IEJ4	Q8iej4 plasmodium

ALIGNMENTS

RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.

ID Q9AHT5; AC Q9AHT5; DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Serine protease (fragment).

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; PubMed=11179332;

RA Wizenann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,

RA Gayle A., Brewh Y.A., Walsh W., Warren P., Lathigra R., Hanson M.,

RA Langermann S., Johnson S., Koenig S.;

RT "Use of a Whole Genome Approach To Identify Vaccine Molecules

RT Affording Protection against Streptococcus pneumoniae Infection.";

RL Infect Immun. 69:1593-1598(2001).

CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY

CC AN AMIDE BOND (BY SIMILARITY).

CC EMBL; AF291699; AAK19159.1; -.

CC HSSP; P00782; 2SBT.

DR MEROPS; S08.064; -.

DR InterPro; IPR001899; Gram_pos_anchor.

DR InterPro; IPR006192; LPXTG.

DR InterPro; IPR003137; PA.

DR InterPro; IPR00209; Peptidase_S8.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00746; Gram_pos_anchor; 1.

DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase_S8; 2.

DR PRINTS; PR00723; SUBTILISIN.

DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.

```

DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 51798756B960A6A CRC64;

Query Match
Best Local Similarity 100.0%; Score 4026; DB 2; Length 2119;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKESLKKDTTGVHHHQEENESIKESSTFIDRNISTIRDFENK 60
DB 1313 KLGEIAESKFNGLNGKESLKKDTTGVHHHQEENESIKESSTFIDRNISTIRDFENK 1372

QY 61 DLKLLIKKFRVDDFTSETGKRMEEYDYKDDKGNIIAYDGTDLLEYETEKLEIKSKI 120
DB 1373 DLKLLIKKFRVDDFTSETGKRMEEYDYKDDKGNIIAYDGTDLLEYETEKLEIKSKI 1432

QY 121 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMTFDLYANIND 180
DB 1433 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMTFDLYANIND 1492

QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDLSEN 240
DB 1493 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDLSEN 1552

QY 241 KPDNLTKMESGKIYSDSEKQOVLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 300
DB 1553 KPDNLTKMESGKIYSDSEKQOVLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 1612

QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 360
DB 1613 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 1672

QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 420
DB 1673 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 1732

QY 421 SKSNKIYVNPPEYLRGKISDKGFFNWLVRNESVVDNLIYGLDHIIDNTROFNKLVK 480
DB 1733 SKSNKIYVNPPEYLRGKISDKGFFNWLVRNESVVDNLIYGLDHIIDNTROFNKLVK 1792

QY 481 DGDIMDWKMDYKANGFPDKVTDMDGNVYLQGYSDLNKAKAVGVHVFQFLYDNVKNPEVNI 540
DB 1793 DGDIMDWKMDYKANGFPDKVTDMDGNVYLQGYSDLNKAKAVGVHVFQFLYDNVKNPEVNI 1852

QY 541 PKGNTSIEYADGKSVVFNINDKNNNGFDGEIOEHIYINGKEYTSENDIKQIIDKTLNIK 600
DB 1853 PKGNTSIEYADGKSVVFNINDKNNNGFDGEIOEHIYINGKEYTSENDIKQIIDKTLNIK 1912

QY 601 IVVKDFARNTVKEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSBEEDFILPVYKGE 660
DB 1913 IVVKDFARNTVKEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSBEEDFILPVYKGE 1972

QY 661 LEKGYQFDGWEISGFGKDGAGVIVNLKSDTFIKVPVKIIEEKEENKPTFDVSKKON 720
DB 1973 LEKGYQFDGWEISGFGKDGAGVIVNLKSDTFIKVPVKIIEEKEENKPTFDVSKKON 2032

QY 721 PQVNHSQLNESHKEDLQREHSQKSDSKDVTATVLDKNNISSKSTTNPNK 773
DB 2033 PQVNHSQLNESHKEDLQREHSQKSDSKDVTATVLDKNNISSKSTTNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

```

```

DE Serine protease, subtilase family.
GN SP0641
OC Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Purkerson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri S., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAM; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match
Best Local Similarity 100.0%; Score 4026; DB 16; Length 2140;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKESLKKDTTGVHHHQEENESIKESSTFIDRNISTIRDFENK 60
DB 1334 KLGEIAESKFNGLNGKESLKKDTTGVHHHQEENESIKESSTFIDRNISTIRDFENK 1393

QY 61 DLKLLIKKFRVDDFTSETGKRMEEYDYKDDKGNIIAYDGTDLLEYETEKLEIKSKI 120
DB 1394 DLKLLIKKFRVDDFTSETGKRMEEYDYKDDKGNIIAYDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMTFDLYANIND 180
DB 1454 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMTFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDLSEN 240
DB 1514 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDLSEN 1573

QY 241 KPDNLTKMESGKIYSDSEKQOVLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 300
DB 1574 KPDNLTKMESGKIYSDSEKQOVLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 1633

QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 360
DB 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 1693

QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 420
DB 1694 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 1753

```

QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 480
 DB 1754 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 1813
 QY 481 DGDIMDGMKDYKANGFPDVKVTDGNGVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 540
 DB 1814 DGDIMDGMKDYKANGFPDVKVTDGNGVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 1873
 QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNKK 600
 DB 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNKK 1933
 QY 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 660
 DB 1934 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 1993
 QY 661 LEKGQYFDGWEISGPEGKDGAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKDN 720
 DB 1994 LEKGQYFDGWEISGPEGKDGAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKDN 2053
 QY 721 POVNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
 DB 2054 POVNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 3

Q9S4M8 ID Q9S4M8 PRELIMINARY; PRT; 2144 AA.
 AC Q9S4M8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA precursor.
 GN PRtA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN (1)
 RC SEQUENCE FROM N.A.
 RC STRAIN=3.B;
 RA Bethel G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;
 RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae."
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).
 CC EMBL; AF127143; AAD48399.1; -;
 DR HSP; P00782; 2SST.
 DR MEROPS; S08.064; -;
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PR00723; SUBTILASIN.
 DR TIGRfams; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 KW Cell wall; Peptidoglycan-anchor; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE PRtA.
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;

Query Match

94.2%; Score 3793.5; DB 2; Length 2144;

Best Local Similarity 94.7%; Pred. No. 1.5e-144;
 Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;
 QY 1 KLGEIAESKFKNLGNKGEGSLKDDTTGVHHHQQEENEESEIKESKSFITDRNISTIRDFENK 60
 DB 1339 KLGEIPESKFKNLKXVKKDDSLNKETAENVNLLVDNQSIGSLFNHKTISTIRDFENK 1398
 QY 61 DLKULIKKKPREVDFTSETGRMBEYDYKYDDKGNIIAYDDGTDLLEYTEKLEIKSKI 120
 DB 1399 DLKULIKKKYQEDDFVN-GGTRITVERDYKYDDKGNIIAYDDGTDLLEYTEKLEIKSKI 1457
 QY 121 YGVLTSPSKDGHPEILGKISNWSKNAKYVYGNKYSEIKATKYDFHSTKMTFDLVANIND 180
 DB 1458 YGVLTSPSKDGHPEILGKISNWSKNAKYVYGNKYSEIKATKYDFHSTKMTFDLVANIND 1517
 QY 181 IVDGLAFAGDMELFVKDNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSLN 240
 DB 1518 IVDGLAFAGDMELFVKDNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSLN 1577
 QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNILIRGYALKVTTYNPNGTMDLEGNVYSKEDI 300
 DB 1578 KPDNLTKMESGKIYSDSEKQVLLKDNILIRGYALKVTTYNPNGTMDLEGNVYSKEDI 1637
 QY 301 AKIQANPNLRALSETTIYADSRNVEDGSRSTOSVLSALDGENIIRYQVTFPMNDKGEA 360
 DB 1638 AKIQANPNLRALSETTIYADSRNVEDGSRSTOSVLSALDGENIIRYQVTFPMNDKGEA 1697
 QY 361 IDKGNLVTDSKLVLFKDDKEYTGEDKFNVEALIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
 DB 1698 IDKGNLVTDSKLVLFKDDKEYTGEDKFNVEALIKEDGSMFLFIDTKPVNLSMDKNYFNP 1757
 QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 480
 DB 1758 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 1817
 QY 481 DGDIMDGMKDYKANGFPDVKVTDGNGVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 540
 DB 1818 DGDIMDGMKDYKANGFPDVKVTDGNGVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 1877
 QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNKK 600
 DB 1878 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNKK 1937
 QY 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 660
 DB 1938 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 1997
 QY 661 LEKGQYFDGWEISGPEGKDGAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKDN 720
 DB 1998 LEKGQYFDGWEISGPEGKDGAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKDN 2057
 QY 721 POVNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
 DB 2058 POVNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2110
 RESULT 4
 Q8DQ7 ID Q8DQ7 PRELIMINARY; PRT; 2144 AA.
 AC Q8DQ7;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
 GN PRtA OR SPR0561.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,


```
Db 692 DLNKRKSNIDIQVNNNNNNNDTTFYDAMEYSGNDKEQVKNVQED-HVINGVESKNKN 750
QY 681 -----AGYVNLSDTFI-----KPVF-- 697
Db 751 MOTNCSSNNKYNINDRWNHLKYEYDMKPGVDNIINIKENELVLQKNEKKDIFNK 810
QY 698 ----KKIEKKEEENK--PTFDVSKKQDNPQVNH-----SOLNESH-----RKEDLQRE 740
Db 811 SHGNEKIEPKDFNNKVNKVTNLKKNVNDIHEDEVKSVNIGNEIDENRKKIKKENILNS 870
QY 741 EHSKSDSKD-VTATVLDKNNISSKSTNNPN 772
Db 871 INQEKDIGKNIIKNTSYSKNLLSKYAEKPH 903

RESULT 6
Q81L45
ID Q81L45 PRELIMINARY; PRT; 3504 AA.
AC Q81L45;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf14_0404.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaigya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL; AE014822; AAN37017.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3504 AA; 408303 MW; B8454D48D5BE4F0 CRC64;

Query Match 6.8%; Score 264; DB 5; Length 3504;
Best Local Similarity 21.2%; Pred. No. 0.018;
Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;

QY 11 KNLGNGKGLKDTTGVVHHQEN--EESIKEK-----SSFTIDRNISTIRDENKDL 62
Db 1432 KNSHENSENIMTEYGNKNSPOENIDDIIPKNGENKNSQNSDHNIMTKNGENKNS 1491
QY 63 KKLKKKKFREVDFTSETGRMBEYDYKYDDK----GNIIAYDDGTDLEYTEKLDKIS 118
Db 1492 QQ-----NSDHNIMTEYGNKNSQNTDDNIMTEYGNKNSKNEKKEEDIAS 1539
QY 119 KIYGV----LSPSKD-GHFEILGKINSVKNKAVYVG-----NNYKSIKATKYDF--- 165
Db 1540 --YEIDKNRISHENDQEHFTPEBSRINKGFBHTNVYNSVGDNEEKGISVKNISEDIIIPD 1597
QY 166 -HSKWTFTFLYANINDIVDLAPAGDMFLPVKNDQKAEIKIRMEPIKETKSE---YP 221
Db 1598 GKGKNIQDDIILEN-----GENKNFEENIEEDKISDTQ-KSKISHENAEHGFTP 1647
QY 222 YVS-----SYGVNIELGEGSLSKPKDNLTKMESGRYI-----SDSEKQOYL 264
Db 1648 YESGKNKISDENDVEYNSDINTNKDQE--EVESKEIFETNDINKNGHISSDNNKINKMK 1705
QY 265 KONILRKGVALKVYTNYPGKTD-MLEGNGVYSKEDIKIQKANPNLR-----ALSETTI 318
```

```
Db 1706 QNNIILNESQDKNIQDVHN--KLDKILKNHHTVSDSELEKIKEENGTRKNKKNINNEEKI 1763
QY 319 YADSRNV-----EDGRSTQSVLMSALDFNRIYQVFTFKNN----- 355
Db 1764 BEEKENVKNDETIIGKKEENTESDDLKIQKISNENINKNILYTDNYNKDKSYNAQGGTHG 1823
QY 356 -----DKGEAID-----KGNLV----- 368
Db 1824 ENDETTNGTNIISGDLKDNVKIQDVISKENTLOENKEDIIPSVTTINNSLGDYKVKENLS 1883
QY 369 -TDSSKLVLFQKDDKEYTGEDKFNV-----BAIKEDGSMFLFDTK-----PVNLS- 412
Db 1884 PEDIKKMEVAHKNIQNIITSEDELGTQKDNERNKEDKSPNGVEENHQENDKIIGEVLNSN 1943
QY 413 MDKNYFNPSKSNKIYVRNPEFYLRGK-ISKDGGFNWELR--VNESVVDNY-----L 460
Db 1944 MNMESNIGNSDTI---NOHLLNEGKNIIHKGNVNSETNMTNNSGTQNIISNEQFEKNI 2000
QY 461 IYGDLIHIDNTRDFNKLNVKQGDIMDMGMYKANGFPDKVTDMDGNVYVLTQGYSLNAK 520
Db 2001 IRGD-DIKDKMNVKIEDETCN-----NIKINKYNDNAKILNELIINKOQTQSDAD 2052
QY 521 AVGVHYQFLYDNVK-PEVNIDPKGNTSLEYADGKSVFVNINDKRNNGPDGEIQEOHIYIN 579
Db 2053 DISTNGSKMDQIENKNNENIHKNNVVE-----KDKI--SNDKEINIVPPE-HKEHDIIS 2105
QY 580 KGEYTSFNDIKQIIDKTLNLIKIVKDFARNTTVKFELNKDTCVSELKPHRVTVTIQNG 639
Db 2106 DNKKKEFDNVLEIPKGNH-----LDDKETITEQVEKSIQODKSMENNVT--NDG 2157
QY 640 KEMSTVISEEDFILPVYKGELEKGYQDFGWEISGFEGKKDAGYVINLSK----- 689
Db 2158 KDIIH---IQEEDI-----KGINNVNNDKHSKKNLHIDEPNKYVEEKEIKKH 2204
QY 690 ----DTFIKPVKKEIEKEEENKPT-----FDVSKKKNQPVNHSOLNESHKEDLQREE 741
Db 2205 EIADHDIKKFKBKEIQENOSNKNPSENENILVDVNAQDDK-----NISKLTDNLHDOE 2256
QY 742 HSQKSDSTKDVATVTLVDKNNISSKS 766
Db 2257 KGTNDSVVE-----HNVSDDKT 2273

RESULT 7
ID Q813P4 PRELIMINARY; PRT; 1777 AA.
AC Q813P4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFE1095W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (S88-2002) to the EMBL/Genbank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Clark R.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Keshornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
```

RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Whitin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929353; CAD51583.1; --
KW Hypothetical protein.
SQ SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;
Query Match 6.3%; Score 255; DB 5; Length 1777;
Best Local Similarity 20.7%; Pred. No. 0.019;
Matches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;
QY 4 ETAEKFKNLGKESLKKDTTGVVHHQENE-----ESTKESSTIDRNISTIRDF 57
DB 977 ELLDDKKKL--DEENELDDKK--KKLDEENELDDKKLDEENELDDKKKL--DE 1030
QY 58 ENK---DLKKLI-----KKKPREVDFTSETGKMEYDYKDDKGNIIAYDDGT 104
DB 1031 ENELDDKKKLDEENELDDKKLDEENELDEERKKKMDENILLDEKKKEIVNDHNT 1090
QY 105 DLEYETEKLDEIKSKI---YGVLSPS-KDGHF-----EILGKISNVSKNAK----- 146
DB 1091 FIOTEHNLNKKTKLOEDYNIIEAELEKWHKSKILLEETKIKIENDENDIKRKQSQIEN 1150
QY 147 VYGGNNYKSIETKATYDFHSTMTPTDLYANINDIVDGLAFAGDMRLFKVNDQKAEIK 206
DB 1151 IYRRNSMMDININSYKSNVKTTFNFDNIEE----- 1184
QY 207 IRMPEKIKETKEEYPVSYGNGVIELGEGDLKSKPDNLTKMESGKIYSDSEKQVLLKD 266
DB 1185 ----EDKITQKNTIYTDISNM-----LTKNKSSTYSINSEKDIINEQEYISNKN 1234
QY 267 NI-----ILRGYALKVTTNPGKTMLEGNGVYSK-----EDIAKIQK-ANPN 309
DB 1235 NTFNNDIEKDLTNMKEKNFNINISYKQCSNIYDINDSNKFLMDTESIIKQINDINEK 1294
QY 310 LRALS---ETTYADSRNVEDGRSTQSLM--SALDGFNIIRYQVTFPMNDKGEAIDKGN 366
DB 1295 VKLLKQREEIFYBEKKNFKEKELHLLKENVLNKMWIIKDR--ENELNKKBEELKMKKE 1352
QY 367 LVTDSSKLVL-----GKDKKEY-----TCEDKFNVEAIKEDGSMFLFIDTKPNLSM 413
DB 1353 ILLSKENIINYSLNSSKINKKLEYNKLEVDYTKGGINMTSKFSKSHL--SDTN--NIOP 1409
QY 414 DKNYFNPSKNIYVRNPEFLRGKISDKGGFWELRV-----NESVDNLIY-G 463
DB 1410 FKNDIN---NDKLYEDNISY--GNLSHDNNMYDNNLSCDNNMSHDNNWARDNNICHN 1465
QY 464 DLHIDNTRDFNIKLV-----KGDIMDMGMDYKANGPDKVTDMDGVY----- 509
DB 1466 NISYDNPSTNTHMKLSELENNENKKNIDY---YDKNFDVRNIIINHNFNISKI 1521
QY 510 ----LQGYSDLNKAVGVHYQFYDYNKPEVNIDPKGNTSIEYADGSKSVFVINDKRNNG 566
DB 1522 SNDISGNDMDNE-----LHDNL-----NNSLKEIEMYKSLLSRSDSEINT 1562
QY 567 FDEIEQEQ-----HIYINGK-----EY-TSFNDIKQIIDKTLN 598
DB 1563 LKLKIEKQTEKWKQNNINOLNDISDNNIYINGNISNGVNYDTSMHVELPERKFLN 1622
QY 599 IKIVKDPARNTVTKFILNKOTG---EVSLEKPHRVTVTIQNGKMSSTIVSEEDFI-- 653
DB 1623 LS-EVED---NEKMSILSNELSVLYKEINNIK-----EYNINVLKNEFIGN 1667
QY 654 -LPVYGELEKGYQFGWELSGEGKADAGYVNLNSKDTFIKVPVFKIEKKEEKNKPTF 712
DB 1668 LLNLFNLDLRNNYRL---KENYYEKEVHKSQIIISERDFIKELQNTINEKKLKE----- 1719
QY 713 DVSKKDNPOVNSQLNESHK---KEDLQREHSQKSDSTKDVATVLDKNNISSKST 767

DB 1720 -ISYKMKMLKWN--QINDTYKLKNRSLSTVELLQK--DIKFLNEDVLKMKEMVTILST 1772
RESULT 8
QYVTV6 PRELIMINARY; PRT; 1127 AA.
AC QYVTV6;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DE 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE ORF MSV156 hypothetical protein.
GN MSV156.
OS Melanoplus sanguinipes entomopoxvirus (MsEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RX MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
RL J. Virol. 73:533-552(1999).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063866; AAC97677.1; --
KW Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F185DA1D5A3FE7D1 CRC64;
Query Match 6.3%; Score 253.5; DB 12; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.013;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
QY 7 ESKFNLGNGKSGSLKKDTTGVVHHQENEESIK--EKSFTID-RNISTIRDFENKOLK 63
DB 196 EIEFNIDNVQKEINKKQDELNKLDESKEIKKQEEELNKTIDKQBELIKKLNDKEIN 255
QY 64 KLKKKPREVDFTSETGKME-----EYDYKDDKGNIIAYDDGT--DLEYETE 111
DB 256 FNIDEQKLLDQINSKINTLNENIKGVMNLYETKNKISNLQNEILNKDSTIKSLDEKQ 315
QY 112 KLDEIKSKIYGVLSKDGHPILGKISNVSK-----NAKVVYGNVYSKIEIKAT 161
DB 316 LLDELKDNNTISLYNKS-----TKITNIQQLLESSLTFNNANI-----NINELSKIK 367
QY 162 KYDFHSKTFDLYANINDIVDGLAFAGDMRLFVK--DNDOKKA-EIKIRMPKIKET-- 216
DB 368 LFDNDIQKLNNDITEQNKKITD--FPNNSTRIFKEKLDTEYKKIDDIKNNLQKLESYK 425
QY 217 ----KSEY---PYVSYGNGVIELGEGDLKSKPDNLTKMESGKIYSDSEKQVLLKNDII 269
DB 426 KIDEQTEYVKNKINKEYNDIIEELKNNLQKLEENKKIDEQTEYVKNKINKEY--NDIIE 483
QY 270 LRKGVALKVTTPNG-----KIDMLEGNGVYSKEDIKQKAPNLRALSETTIYAD 321
DB 484 LKNNLQKLEENKINDKLTKLNDIESNTELFKNLISDFKDKSREIAKLN--TEYEQ 541
QY 322 SRN--VEDGRSTQSVL-----MSALDGFNIIRYQVTFKMDNGKSAIDKGNLVTDSK 373
DB 542 LRKDLLENKNTNMLKSLSDNKLSSLE-----QLYDSKKNIL-DGIDKIYNSLAKEN- 592
QY 374 LVLFGKDDKEYTGEDKFNVEAIKED---GSMFLFIDTKPNVLSMDKQVFN---PSKNNKIY 427
DB 593 ----DKIDEYFSNIEKFIYVNIENKFNIGNLSIINKIINNDFQKEYINSKIDSKSNELS 648
QY 428 VRNPEFY-----LRKISDKGFGNFWELRVNESVDNLYIGDLHIHNTRD 472
DB 649 TWFDDIFNAKQIASITNNIENISNKKID-----LNEFTISNEDSKEL-LDEIRK 698
QY 473 FNIKLVKGDITMDGMDKYKANGPDKVTDMDGNV-YLQGTGYSDLNKAKA-----V 522


```
Db 699 YKQFDP-KTKDAMNTEVKGFE-NLQKDDISIKSNELNLTAYDIINTKANDLDDKLANN 756
Qy 523 GVHYQFLYDNVXK------EWNIDPKGNTSIEYAD-GKSVVFNIDKRNNGFDGEIQEHI 576
Db 757 GSEFKLYNNASDLLLOTIOKNDKVKQLNEYLEKKNQOSIENDIVN-----805
Qy 577 YINGKBYTGFNDIKIIDKTLNKKIVVQDFARNTTVKEFELNKDGEVSE---LKPFRVT 633
Db 806 FI--KELIKFNTE--TNKSLN-ELLTND---DINDKIFKLYKELNKKISTNNLLKIYK-- 855
Qy 634 VTIONKEMSSITVSEDFILPVYKGELEKGYQDFGWEISGPEGKDGAGVINLSKDTF- 692
Db 856 NEIDNVNKLKSVIENLQFNSFLSEFNQG-----SITSH-----INFLNLTIA 900
Qy 693 -IKPVFKKIEEKEENKPTFDVSKKKNQPNVHNSQLNESHKEDLQREHSOKSDSTKD 751
Db 901 GINDVLNKLNLKIMADTTTRGDINR-----DEIKNQISSENIKSKOFNEKNE--KD 950
Qy 752 VTATV-----LDKNISSKSTNN 770
Db 951 LKKLISFNDKLNKYNISAGVTEYN 974
```

RESULT 9

```
Q81121 ID Q81121 PRELIMINARY; PRT; 2227 AA.
AC Q81121;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0354.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.B., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J.J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014841; AA035938.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2227 AA; 262840 MW; 6D5D4BEEF643339C CRC64;
```

Query Match 6.3%; Score 252.5; DB 5; Length 2227;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;

```
Qy 10 FKNLNGKE-----GSLKDDTTGVH-----HHQNESSIKK 42
Db 690 YNNNNNNKCTCTSIKSHENKYPFKSHVSNYMKNTNHLPHRNAITSNNRNEEYK 749
Qy 43 SSFTIDRNTST-----IRDPENK-----DLKKLI-----KKKPREVDFT-----SE 79
Db 750 EK---DRNTGNNNLYEVNNSCIPPLKKMIPIDGVNRKNGINKLNNTVNTQRTSSVS 806
Qy 80 TGRMEEDY-----KYDDKGNIIAYDDGTDLEVEYKLEIKSKIYGVLSPSK 128
Db 807 TKNDIENSFDMPINGIRESKYISNNNI---NGSIGFNSKLDNYHQSNV-----858
Qy 129 DGHFEILGISVSKNKKVYGNKYKSIKATKYDFHSKWTTF-----DLYANIN---179
```

```
Db 859 -----NESYPLKMMKNY-----IEHNYDDKNIFLVKNYEDTYSNIHGI 900
Qy 180 -----DIVDGLAFAGDME-----LPVKDN-----DOKK-----202
Db 901 HENSMKYNVNLKACFFHGRHKNQKMYTEENLNINQKNYSHYHNGTVLKPVTNN 960
Qy 203 -----AETIKRMPKIKETKS-----EYPYVSSYGNVIELGEGDLSKQKPDNLTKMESGKI 253
Db 961 VAVNEFADINLSAQKRLHSLKSMGYEDKSMENYRNKIYNNINNNNNNDN-----NI 1013
Qy 254 YDSEKQOY-----LLKDNIIILKGYAL-----KVTYNPCKTDMLENGYVSKEDIAK---302
Db 1014 YNDNEYCOYNSYCFDHSDLKNMFLNQHNSKLLTHSNKNKSFNGINYESKHLANPEI 1073
Qy 303 -----IOKANP-----N 309
Db 1074 KTFAHNSYPILNOGLINCNPLOCLGYDSNQKNHNVVYIKKNEYLKNIGSIINVLKREG 1133
Qy 310 LRALS-----ETTIYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPMND 356
Db 1134 LRKISTHNGKPFESFNMNDKNVMEGLNIQDNVNNNNKESC-DNIKHMRTKSLNFSVRE 1192
Qy 357 K-GE--AIDKGNLVTSSKLVLFGK-DDKEYTGEDKFNVEAIKED-GSMLFID---TKP 408
Db 1193 SYGEHKSLLDVYQECYVKNKLI--NKVNDKCY--EDNNSYLNEDDNASMQFYETNSNP 1248
Qy 409 VNLSMD---KNYFNPSKSNKIYVRNPEFYLRGKISDKGGFNWE-----LRVNESV-----455
Db 1249 YIVDQENNMKNYV-----NVLNNNSNYVDSKNYDKSKENAEKSDDDLNNENIHTLKD 1304
Qy 456 -----VDNLYIYDGLHIDNTRDFNFKLVNVDGDIWDV--GMKDYKANGFPDKV-TMDMG 506
Db 1305 OKKKIQNNNEFISEQADIENTRNSQEVVEKEHEPL-WVINASNEEKSVEELIYSDMSS 1363
Qy 507 NYLQGYSDLNKAVGVHYQFLYDNVRPEVNIIDPKGNTSIYADGKSVVFINDKRNNG 566
Db 1364 NRVTNKNYSMDMNVVLLNEDNLLTTEKYKVQLEKENKMDMYETVEENINTIKTENTND 1423
Qy 567 FDEIQEQHIYINGKEYTSFNDIKQIIDLTKIKVVKDFARNT--TVK--EFILNKDTG 622
Db 1424 INEVR-----NEOKRRESININD-----TNINHIIIDYPNDTYNFIDIECVHNNEN 1472
Qy 623 EYSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQDFGWEISGPEGKDGAG 682
Db 1473 MYNSIEQVTFYHDTNRNHLVDK--NNQNF-----FEIEGLNELNFEKK--V 1516
Qy 683 YVINLSKD-----TFIKPVFKKIEEKEENKPTF-----DVSKKKDN 720
Db 1517 YIENNTKDDHKGDSKTSNLTSLRNTICKSENHNEKNENTYVVRKGEKGRKRVSMKKRN 1576
Qy 721 POVN-HSQLN-----ESHKEDLQREHSOKS 746
Db 1577 EKLNENYNNIYDKNDNHRQNDITKKENDEEN 1609
```

RESULT 10

```
Q81H4 ID Q81H4 PRELIMINARY; PRT; 2849 AA.
AC Q81H4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0392.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., Berriman M., Hyman R.W.,
```

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalloo S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.,
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum";
RL Nature 419:498-511 (2002).
DR EMBL; AE014841; AAN35975.1; --
KW Hypothetical protein.
SQ SEQUENCE 2849 AA; 335916 MW; B5515D173D96B813 CRC64;

Query Match 6.2%; Score 251.5; DB 5; Length 2849;
Best Local Similarity 19.1%; Pred. No. 0.046;
Matches 178; Conservative 162; Mismatches 361; Indels 231; Gaps 42;

QY 8 SKFNLGKSGSLKDTGTGVEHHQENESIEKSSFTIDRNISTIRDPEN--KDLKK- 64
DB 295 SYNNNYNNNNNN--NNNSANVTNKRDKKYKSVSFANQENHSYKFKQDKPSLKNK 353
QY 65 ---LIKK-KFREVDDFTSETGRMEERYDYKDDKGNIIAYD---DGTD-LEYETEKLDE 115
DB 354 KDTILKKIYEDELYKQYDQMYEGEYDSDE---LYDEMCEDSDDKIYYEDEKFN 409
QY 116 IKSIVGLSPSGKHFEILG--KISNVSKNAKVYGNVYKSIETIKATYDFHSKTMTPD 173
DB 410 MVNK-----KVAKDINFEMWNSKVKISKDINL-----KYKSDNRKMSNDSHVKS--SEN 459
QY 174 LVANINDIVDGLAFAGDMRLFYKNDQKAEIK-----TRMP-----EKIKETSEYP-- 221
DB 460 ILNNNKILNRVLDPKDISYVEEYQKRKEQAQNYHNVENPNYTDKNINDKENYPGD 519
QY 222 -----YVSSYGVNI-----ELGEGDLKKNKPNL-----TKMESGKI 253
DB 520 NINLNNNIRSNNTTVYANSTLSNEKNMNLNSNISDEVSSINKENLLNTNKNSENIT 579
QY 254 YSDSEKQOYLLKDNILKRGYALKVTTYNPGKT-----DMLBEG-----VYSKEDI 300
DB 580 IQNCDDNNKLVNDIVIDQEFLIK--NYNVNKNKDYSTYIDSLKNNNFIDKIYTDNDV 637
QY 301 AKIQ-----KANPNLRALSETTIYADSRNVED-----GRSTQSVLMSALDGF 342
DB 638 INTESMGIYNETNMKNKEISEPCIHNTNINIEYLITNNDMTNMGVNSNDILNKNIEKT 697
QY 343 NIIR-----YQVFT-----FKMNDKGEAIDKGNLVTDSSKLVLFGKD--DKEYTGEDKEN 391
DB 698 NFENESKLYDVYDMINDYKKNKEETINKIOEKCVDKVMYDFINNIDKETTINLDM-- 755
QY 392 VEAIAKEDGSMFLFDTKPNVLSMD--KYNFNPSSKNKIYV----- 428
DB 756 -----DKKHLVNLNNEKNSTCDDKDTYLEMSKKEKINTFLKYLKMDVNSLSHLFQYFV 809
QY 429 -----RNPEFYLRG---KLSDKGGFWELRVNESVVDNVLIIYGLDLHIDNTRFNK 476
DB 810 DREKDEMKKQLQLLIGDGEKKQMEFNKYKINQNTQTLNKLGHESVOTNNEKKIQIE 869
QY 477 LNVKGDGIMDGMKDYKANGFPDKVTDMDGNVYLGQYSDLNKAVGVHVFYQFLDMVKPE 536
DB 870 -NIQTDIKDTIKTLIKDMINKTISIDSVFF-----KSLSKDSYDLNKNKED 918
QY 537 VNIDPKGNTSIEYADGKVVFNINDKRNNGFDGEIOEHYINGKBYTSFNDIKQIIDKT 596
DB 919 IK---KNDIT-----YTKQEDMENKVD-----VTNVESNTETVNEVQKIND-- 956
QY 597 LNIKIVK-----DFARNITVKE-----FILNKDTGVESELPHVTVTIQNGKMSST 645
DB 957 LKVKILLEIKGCVYDNTSNNNYNDDEQTALILMDKNEYSKEMDYVYNLIDENRILSK 1016
QY 646 IVSEEDFILPVYKGELEKGYQFDGWIISGPEGKDKAGYVINL-----S 688

DB 1017 LNDEENKSNSEYKNNKNSMW-----TVTFASLKSFEREMNLLKSHNRLRRRIKLYES 1072
QY 689 KDTFKPKPFKKEEKEENK---PTFDVSKKNDPQVNHSHOLNESHKEDLQRE--- 741
DB 1073 RDR--IKNEYIKMEKUKESQDRFIATERHIEKHNELNLSKKNED-MKYDLKKNKIKI 1130
QY 742 -HSQKSDSTKQVDTATVLDKNNISSKSTTNNPN 772
DB 1131 ALESQIDNNLNMISNNMEKNNNNNNNNNN 1162

RESULT 11
Q812V4
ID Q812V4 PRELIMINARY; PRT; 3381 AA.
AC Q812V4;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI0975C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berrian M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jafels K., James K.D., Johnson D., Kethornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531 (2002).
DR EMBL; AL929157; CAD51881.1; --
KW Hypothetical protein.
SQ SEQUENCE 3381 AA; 396148 MW; 0DAB0AFAB2EDB88 CRC64;

Query Match 6.2%; Score 251.5; DB 5; Length 3381;
Best Local Similarity 19.9%; Pred. No. 0.055;
Matches 193; Conservative 150; Mismatches 354; Indels 275; Gaps 45;

QY 18 EGSLLKDT-----TGVEHHQENESIEKSSFTIDRNISTIRDPENKDLKK----- 64
DB 801 ENNMKNDFNNKNNTINDEQNDSEIYNSKQDPVSCDDSSV---NKDIRKNYEGIP 856
QY 65 LIKKKPREV-----DDFTSETGK-----RMEYDYKYDD-----KGNIIAY 100
DB 857 VVEVQDEYYENEQNDDEDDDEDDDEDDDEDDDEDDDEDDDEDDDEDDDEDDDEDD 916
QY 101 DDGDTLEYETEKLDEIKSIYG-----VLT-----SPSK 128
DB 917 DDDTNNYEDEB-SFLSTNIFYGKKNNSDNIHNKIIINKVSKDNIVLKSLSKSSSS 975
QY 129 DGHFEILGKISNVKNAKVYGNVYKSIETK-----AT 161
DB 976 DG-----IKKSVSFKYVYDLAKNVEPMNFKSFDDITMSEMDDQDNINVTIPEHTFSQ 1031
QY 162 KYDFHSKTMTPDLYANINDIVDGLAFAGDMRLFYKNDQKAEIK-----NDOKKAEIKI- 207
DB 1032 NKKIYKDMNNTSNTDLTINNYESTFEMKNLEKKNQKDLNLSHKNQKEVIQIN 1091
QY 208 -----RMPEKIKETKSEYPYVSSYGVNIEGEGDLKKNKPNL-----TKMESGKI 253

Db	1092	HLITNMKNNTVLKEELKNNNNHHYEVKK--NINE-----DICNDVNNLGTDTSTQKNFGVT	1144
Qy	254	YS-DSEKQOYLLKDNIIILRGY-----ALKVTTNPGKTDMLGN-----GVYSKEDI	303
Db	1146	PSLSTKNNFGVTPSLDSTKNNFGVTPSLDSTKNNFGVTPSLDSTKNNFGVTHNLDSTKN	1205
Qy	304	Q-KANPNLRALSETTIVADSRNVEDGRSTQSV-----LMSALDGP-----NIIRY	347
Db	1206	NFGVTPNLSDSTKNN--FGVTHNLDSTKNNFGVTPSLDSTKNNFGVTPSLDSTIIKFSDEEK	1263
Qy	348	QVFTFKWDKGEAIDKGNLVTDSKVLVF-----GKDDKEYTCEDKFNVEAIKEDGSM	402
Db	1264	EQYNTKLNTTNYNTKESNKFIEBEKKVVISLDPNKNKNNNNIN--NYEEIHKGOYIL	1322
Qy	403	-----FIDTTPKVLNS-----MDKNYFNPSKSNKIYVRNPEFYLRG	437
Db	1323	RNYRTNENMEBEOKKNSFFLINQININSINIKGHMDELGLIHNNNNKYVNDESKTYFI	1382
Qy	438	KISD-KGGFNWEIWRVNSVVDNYLIYGDLDHIDNTRDNIKLNVKGDIDMDGWMKDYKANG	496
Db	1383	NIIEGKQDNEKENHINEPNVKNNFI-----HIENVK-----NKKKIYIYTESFMKNNNNN	1432
Qy	497	FPDKVTMDGNVYLQTCYS-D-LNAKAVGVHYQFLYDNVKEPVNIDPKGNTSIEYAD--G	552
Db	1433	IYD-----NNNSNNYIYDNYTNMKNKIKTCTCTKYTKNVLK--NRHKTINYNDYVNIIRG	1487
Qy	553	KSVPV---NINDKNNNGFDGEIOBH-----IYNGKEYTSFNDIKQIIDKTLNIKI	601
Db	1488	KNIKFVLITNCNYKTN-----VLKVKHSMKTTKNNVILQNKY-----DLLKLLKNDHNKL	1539
Qy	602	VVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNG-----KEMSSITVISEEDFILPVY	657
Db	1540	LYNHL-----YLIQREAHILLRVNVKYGVLYNKGHNHSRKQNNNIKKKRSISFSS	1591
Qy	658	KGELEKYQDPGEWISGF-----EGKQDAGYVINLSKDTFIPKVPFKIBEKK	704
Db	1592	LNQODNSN--DNSDVSNFIIKNITHYINDIEKNKSNEYTKVNDNDNNSNMLKKEKKK	1649
Qy	705	EEENKPTFDYSKKKNPQVNHSQLNESHREDLQREBSKOSKSTKDVATVLDKNN---	761
Db	1650	NTKDKNTKDKNTKDKN-----TKDKNTKDKNTKDKNTKDKNTKDKNTTKKNNSYM	1701
Qy	762	ISSKSTTTNNPNK	773
Db	1702	VSDNYILYNDNK	1713
RESULT 12			
Q81E65			
ID	Q81E65	PRELIMINARY;	PRT; 3519 AA.
AC	Q81E65		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Hypothetical malaria antigen.		
GN	M26-32-10.		
OS	Plasmodium falciparum (isolate 3D7).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI TaxID=36329;		

RESULT 12

```

Qy 610 -----TTVKEFILNKDTGEVSELKPHRVTTIQNGKEMST-----IVSEEDF----- 652
Db 1336 EDKRNKRKNKKKIYINFSDGIQ-----INNKCKSNSEKCKENYICDET 1385
Qy 653 -----ILPVYKGELEKGYQFDGWEISGFGKKDAGVIN-----LSKDTFIKVPVKKIE 701
Db 1386 KNFTNIIIPKYEIMHKNY-----SKSFNNIKGNQNGEIDETNTIKDEIKDKS 1434
Qy 702 EKKEENKPTFDV-SKKDNQPVNHSQLNSESHREKDLQREHSQKSDSTKDVTATVLVDKN 760
Db 1435 KHEEEYEILHKVDNKKNDQQDDHHNDNNKNDQDDDRHNDNNNNKQ-----NDDH 1489
Qy 761 NISKSTNNPNK 773
Db 1490 PDHHNDNNNNK 1502

RESULT 13
Q8ILU2
ID Q8ILU2 PRELIMINARY; PRT; 1474 AA.
AC Q8ILU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.

```

RESULT 13

RESULT 13	
Q8ILU2	PRELIMINARY; PRT; 1474 AA.
ID	Q8ILU2
AC	Q8ILU2;
DT	01-MAR-2003 (TREMBLrel. 23, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Hypothetical protein.

GN PF14 0152.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=36329;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteaux M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Murgall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL; A5014818; AAN36764.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1474 AA; 178501 MW; 7EBAD49E16B4707C CRC64;

Query Match 6.2%; Score 250; DB 5; Length 1474;
Best Local Similarity 20.1%; Pred. No. 0.025;
Matches 185; Conservative 141; Mismatches 329; Indels 264; Gaps 41;

QY 7 ESKFNLGNGKGLSKKDTTGVVEHHQNEESIKESFTTIDRNFENKDLKLI 66
DB 365 EEEYENIKKEEIKDKNII---HEQEKINIDVSSKNM---VASINDVSSKNMVASI 416

QY 67 -----KKKPREVDDTSET-----GKRM-----EEDYK 90
DB 417 NDVPSKNMVASINDVPSKNMVASINDVPSKNMVASINDVPSKNMVASIKDVPNTNEHQK 476

QY 91 YDDKGNIIAYDDGTDLEYETEKLDETKSK--IYGLVSPSKOGHFEILKISNVSKNAV 147
DB 477 KNEESN---KNGTN--YTKKLLNKNKEEYI---KNHRENH-----KN----- 514

QY 148 YGNNYKSIEIKATKYDFHFSKWTFTDLYANINDIVDGLAFAGDMRFLVKDNDQKAEI-- 205
DB 515 -FGNSYKS-----QNYNNNSLNDKYTSK--IDGTRY-----VYLKDEPNHNSKLIY 559

QY 206 -KIRMEPIKETSEYPPYSSYGNVIEL-----GEG-DLSKNK----- 241
DB 560 NKITL-----KKHYTYEDNYYNNKKLYAYHTFNGSEQTRRGYSGINKNENSKNDLYK 612

QY 242 -----PDNLTKMESGKIYDSEK-----QQYLLKDNILIRKGYALKVTTYNPGKT 286
DB 613 RMHTRDNLNYYNSKIYGRSNKLYGATASKDTKKYRSSRNRVNDKYILRKESYKK--- 669

QY 287 DMEEGNVYSKEDIQIANPNLALSETTIVADSRNVEDGDRSTQSLMSALDGF-NII 345
DB 670 ---EYNNIYDEB-----NKSCKYRNDRDVDRK-----GFRNVD 702

QY 346 RYQVFTFKMDKGEAIDKGNLVTDSKGLVFGKDKKEYTGEDKF-----NVEAIKEDG 399
DB 703 REKAHIYKSDKNEHIREGEGKQREDKYDRKREDKYDRKREDKYDRKREDKYDRKRED 762

QY 400 SMLFIDTKPVLNSMDKNY-----FNPSKSKIYVRNPEFYLRGKISDKGGFNWELR-- 450
DB 763 HDRKKDDK-YDRKRDNNYDRKRDNNYDRKKDDK-YDRKRDNNYDRKKDDKYDKYKDDKYE 820

QY 451 -----VNESVVDNYLYGDLIHIDNTRDFNI-----KLVYKGDGIMDWGMDKYKANGFPD 499
DB 821 HKVEDINEHKVDYDINEHKGNNTNEHKGNNTNEHKGNNTNEHKGN-----GTNEHKGNTND 876

QY 500 -----KVTDMGNYVLQGYSDLN-----AKAVGVHYQFLYDNVKPEVNDIPKGNTSIEY 549
DB 877 HKGNNINEHKGNNTNEHKGNNTNEHKGNNTNEHKGNNTNEHKGNNTNEHKGNNTNEHKGN 936

QY 550 ADGKSVVFNINDKRNNGFDGEIQEHYI-INGKEYTSFNDIKQIIDKTLNLIKIVKDFAR 608

DB 937 YKIKNNHYNKSDSDSNMF---MKGNMYNRYNRYTENSSEKKEVEKNRNGLDKYNNSR 993
QY 609 NTVVKEFILNKDTGEVSELK-----PHRVTVIIONGKEMSSSTIVSEDFILPYKGE 661
DB 994 NTSKKYFNKSKDKNYTHNNKYNNNTDTHKNRYTNKSKS---NYYNEKDEKNKNNIIGTH 1050
QY 662 EKGYDFGWEISGFBGKKGAGVIVNLSKDTFKPKVFKKIEEK----- 703
DB 1051 EK-----DNKD-----LNDINDSNIRNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 1093
QY 704 -----KEENKPTFDVSKKNDPNQVNSHQSOLNESHKREDLQREHSOKS---DSTKDV 753
DB 1094 IDNMGNNKNNNDLMDKNKNDKNKNDKNKNDKNKNDKNKNDKNKNDKNKNDKNKNDKN 1153
QY 754 ATVLDKNNISSKSTNNPN 772
DB 1154 YTLQNNKNNATDPIDKNDN 1172

RESULT 14
O77328
ID O77328 PRELIMINARY; PRT; 1650 AA.
AC O77328;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative serine/threonine protein kinase.
GN PFC0385C, MAL3P3.12.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornaby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum";
RL Nature 400:532-538 (1999).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; Z98547; CAB11112.2; -.
DR HSP; O63450; 1A06.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00200; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1650 AA; 193755 MW; D70FE19C5C640B5A CRC64;

Query Match 6.2%; Score 250; DB 5; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.028;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

QY 20 SLKKDPTGVVEHHQNEESIKESFTTIDRNFENKDLKLIKKKPRE---VDFP 76
DB 394 SLNENYNERKIYKEINKKEYSNKEYV-----HFKQND-DSSIKKNNSSSECLDQ 443

QY 77 TSETGKR--MEEDYKYDDKGNIIAY--DDGTDLE---YETEKLEIKSKIYGLVSPSKD 129
DB 444 KKKTYKTYIIIEQRYNFNDRDNNNAIKDTHHKSGYLYNNIMVSEYKKG--SNNKM 501

QY 130 GHFEILGKIS-----NVSKNAKVYVYGNKYSI--BIKATKYDFHSGTMTFDLYANINDIV 182

```
Db 502 DMEIYNQNTDNIENLNKLYF--DDYEGVDPEKKKKKLDHIIYTOQKEYKNINDIL 560
QY 183 DGLAFAGDRLFKVNDQKAEIKIRMEPIK-----ETKSEYPYVSSVGNVIELGEG-- 235
Db 561 K-----DHLNDKETKEKKNEIEBEEKKNKIEBEEKKNKIEBEEKK 614
QY 236 ----DLGKNKPDNLTRMESKIVSDSEKQOVLKDNII-----LR 271
Db 615 KKIEBEEKKKKI-EMEEBKNKIDDEKNTYANDKIISHIDNVNCNIKIDALLDHEEKK 673
QY 272 KGVALKVTYVNEKTD-----MLEGNGVYSKEDIK-----IQANPNL-- 310
Db 674 KTGHEKEINLYKETKNEYQKMLNDENSIMLEHEKKYNTHQVNNLCTDKMLQKENKILTN 733
QY 311 RALSETTIYADSRNEDGRSTOSVLSALDGFNIIRYQVFTFMNDKGEAIDKGNLVD 370
Db 734 NDKKRTFLLSKSNII-----TSNVLSSKIPG-----TLSTKLNATIKTIKKD---VTD 779
QY 371 SSKLVLF-----KDDKE-----YTGEDKF-NVEA-----IKEDGSMFLDPTK 407
Db 780 NEKKYVVDHRKDNIIKRNKEFINIYKGRNVAANVEIGSEVCNKNVNVKGDNDKMMVENK 839
QY 408 P-----VNLSDMKNYF-----NPSKSNKIYVRNPEF 433
Db 840 QGDDNNMMVENKQGDENNIVVKNIERSESSFMPTHRKNITSTSTDTCAKNKEQIKYPHL 899
QY 434 YLRGKISDKGFFNWEILR-VNESVDVNY-----LIVGDLHIDN-----469
Db 900 Y-SNKNEKDKKSIIFLNKINENIKKYNKDKKIEKISTLEKVFVKQVKNVITNDEKHTSSK 958
QY 470 -TRDFNI-----KLN-----478
Db 959 INDDFNITIDKQKGLNPNPDLNRKIKNETKILEKDKSHMSKIQNNLMKKKTNPPTNNK 1018
QY 479 -VKGDIMDNGKDYKANGFPDKVTDMDGNVL-----QTG 513
Db 1019 GSISSTSISSSTKNFDCGIIIEKNKLSNLKYTCIRKNQNVDSIKLNDKADLYDKCKKTS 1078
QY 514 YSDLNKAVGVHYQFLYDNDVKNPEVNDPKGNTSIEYADGKSVFN-----INDKRNNGPDG 569
Db 1079 FNDINRAAGMMPK-----KRDV-----PNKNMVDTNKGR-VFNPVTLNNYRN----1124
QY 570 EIQEQHIYINGKEYTSFNDIKQIIDKTLNIIKIVVDKFAFARNTTVKPEFILNKOTGEVSELKP 629
Db 1125 -----YIRSNK-----NNVK-----GKKIVLLKE--KQKSLHP 1156
QY 630 HRVTVIQNGKEMSSTIVSEEDFILPVYKGELEKGYQFPGWEISGEGKQDAGYVNLK 689
Db 1157 EGVEADKKLNSYNDKYLIEKDGFKDIINEEMEK-----YKNNK-MKYKI---K 1201
QY 690 DTFIKPVFKKIEEKEENKPTFVSKKDNQVNHSQLNESHREKEDLOREE---HSQKS 746
Db 1202 SNSIPPIIKKIERKSNNDN-----DNFKNDNINSNNNINSNDKCLFLSEKEDRVHLKNN 1258
QY 747 DSTKQDVATVLDKNNISSTNNPNK 773
Db 1259 NIIVNNTWMPKQSNCDNNTISLKNK 1285

RESULT 15
Q81525 ID Q81525 PRELIMINARY; PRT; 5767 AA.
AC Q81525;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN pFL1930W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN=3D7;
RA MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Praser C.M., Barrrell B.;
RA "genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014850; AAN36472.1; -.
KW Hypothetical protein.
SQ SEQUENCE 5767 AA; 691408 MW; E84A340441C67160 CRC64;

Query Match 6.1%; Score 247.5; DB 5; Length 5767;
Best Local Similarity 20.0%; Pred. No. 0.15;
Matches 178; Conservative 144; Mismatches 310; Indels 259; Gaps 39;

QY 1 KLGE---TAESKFKNLGNGKESGLKDKDTGVVHHQHEEESIKES-----43
Db 4259 KLNELLHYEQMNRKNLDLELE-KYKSEDPTHIVKLSRESEELNEKNKILELQQLIETS 4317
QY 44 ---SFTIDRNTISTIRDFENKDLKKLKKKFRVDDFTSETGKRMEEYDYKYDDKGNIIAY 100
Db 4318 YEINIMIDKNKMLIKE-KDKYEQKIDNLSINDYEKEI-KEMNKEKILKTRKIEKMI 4375
QY 101 DDGTDLLEYETEKLDRIKSIYV-----VLSPSKDGHFEILGKISVNSKNAK 146
Db 4376 DDGEDIKKLDLDEAQELILDLEKEKEELNINIELNSNKMDRLDIDVLLSNVDNLNK 4435
QY 147 VYVGNVYSIEIKATYDFHSKTMTFDLYANINDIVDGLAFAGDNRFLFKVNDQKAEIK 206
Db 4436 ENILNEK-----MNEDEQKYNELK 4455
QY 207 TRMPEKIKETKSEYPPVSSYGNVIELGBGLSKNKPDLNLTKMSGKIVSDSKQOVLKLD 266
Db 4456 IAYDQKVHEFNKLYNMLS-----KKNFD-----SENDEQNYTND 4493
QY 267 NIILRGKALKVTTYNPGKTMLEGNGVYSKEDIKIQANPNLRLALSETTIYADSRNVE 326
Db 4494 NVNMOK-----SRNSNNKYDINDFE--NDTIILALNEKNMIESLNEK 4534
QY 327 DGRSTQSVLSALDGFNIIRYQVFTFMNDKGEAIDKGNLVTDSKLVLFKDD-----381
Db 4535 -----IAHVEEENYK---KEELVNYKNNMINELSNKIHIIFEDNLDLI 4574
QY 382 ---KEYTGEDKFNVEAIKE--DGSMLPIDTKPVNLSMDKNYF-----418
Db 4575 SLNEFIKTNFTYQNILDLNNSIDY--KKLLSKILVENQFYKYIISTRWMLFEIKSKS 4632
QY 419 ----NPSKSNKIYVRNP---EFLYLRGKISDKGFFNWEILRVNESVVDNVLIIYDGLHDNT- 470
Db 4633 YSNSNNNNNNIIISNNSNTRESVLRLKKGNF---KRTASVTSN--IYNDLFLSENV 4686
QY 471 -----RDFNIKLVNDGDIIMDGMKDYKANGFPDKVTDMDGNVYLQTGVSDLNA 519
Db 4687 LSAEKENDEIMNSFNKLTNLNVIDDDF-----NSFENICIELN-NLFIEDWNNIFI 4735
QY 520 KAVGVHYQF-----LYDNVKNPEVNIDPKGNTSIEYADGKSVVFNINDKRNNGFDGEIQB- 573
Db 4736 KLNILNENQFNIIILIIHEMKKINKIKNNNNEKLELKSV--KILKRLNISIDLEKR 4793
QY 574 ----QHIYIN-GKE-----YTSFNDIKQIIDKTLN-----IKIVVDFARNTTVKE 614
Db 4794 TEEIEHLNALKCEAKEDILLNQVEDIKKYI-KQLNEQYNEQIILIQNEVELKSLSDQ 4852
QY 615 FILNKDTEVSELKPHRVTVTIQNGKEMSSTIVSEEDFIL-----PVYKGELEK 664
Db [1]
```

Db 4853 LLLKNE---NQIKTEKINKLNSDLNHFNTSLDKSAYQLSELNNEKKKKILLNELEEK 4909
Qy 665 YQFDGWEISGPEGKDGAGYVILNSKDTFIKPVFKKI--EEKKEENKPTFDVSKKD-NP 721
Db 4910 EKIISHLQKDFESNVE--VINELKE-FNEMLINKLKEEYKKEKDNVYD--KNSDHNY 4963
Qy 722 QVNHSQLNESHKEDLQREEHSQKSDSTKDVTATVLDKNISSKSTTNNPN 772
Db 4964 SNNNDNNNNNNSSNRSHNNNNNNNSW-----NNMKSFHKCDIPN 5009

Search completed: January 28, 2004, 13:01:39
Job time : 48 secs